



University of Pennsylvania  
**ScholarlyCommons**

---

Statistics Papers

Wharton Faculty Research

---

6-2013

# Optimal Rates of Convergence for Estimating Toeplitz Covariance Matrices

T. Tony Cai  
*Wharton*

Zhao Ren

Harrison H. Zhou

Follow this and additional works at: [http://repository.upenn.edu/statistics\\_papers](http://repository.upenn.edu/statistics_papers)

 Part of the [Statistics and Probability Commons](#)

---

## Recommended Citation

Cai, T., Ren, Z., & Zhou, H. H. (2013). Optimal Rates of Convergence for Estimating Toeplitz Covariance Matrices. *Probability Theory and Related Fields*, 156 (1), 101-143. <http://dx.doi.org/10.1007/s00440-012-0422-7>

This paper is posted at ScholarlyCommons. [http://repository.upenn.edu/statistics\\_papers/534](http://repository.upenn.edu/statistics_papers/534)  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

# Optimal Rates of Convergence for Estimating Toeplitz Covariance Matrices

## Abstract

Toeplitz covariance matrices are used in the analysis of stationary stochastic processes and a wide range of applications including radar imaging, target detection, speech recognition, and communications systems. In this paper, we consider optimal estimation of large Toeplitz covariance matrices and establish the minimax rate of convergence for two commonly used parameter spaces under the spectral norm. The properties of the tapering and banding estimators are studied in detail and are used to obtain the minimax upper bound. The results also reveal a fundamental difference between the tapering and banding estimators over certain parameter spaces. The minimax lower bound is derived through a novel construction of a more informative experiment for which the minimax lower bound is obtained through an equivalent Gaussian scale model and through a careful selection of a finite collection of least favorable parameters. In addition, optimal rate of convergence for estimating the inverse of a Toeplitz covariance matrix is also established.

## Keywords

banding, covariance matrix, minimax lower bound, optimal rate of convergence, spectral norm, tapering, Toeplitz covariance matrix

## Disciplines

Statistics and Probability

# Optimal Rates of Convergence for Estimating Toeplitz Covariance Matrices

T. Tony Cai<sup>1</sup>, Zhao Ren<sup>2</sup> and Harrison H. Zhou<sup>2</sup>  
University of Pennsylvania and Yale University

## Abstract

Toeplitz covariance matrices are used in the analysis of stationary stochastic processes and a wide range of applications including radar imaging, target detection, speech recognition, and communications systems. In this paper, we consider optimal estimation of large Toeplitz covariance matrices and establish the minimax rate of convergence for two commonly used parameter spaces under the spectral norm. The properties of the tapering and banding estimators are studied in detail and are used to obtain the minimax upper bound. The results also reveal a fundamental difference between the tapering and banding estimators over certain parameter spaces. The minimax lower bound is derived through a novel construction of a more informative experiment for which the minimax lower bound is obtained through an equivalent Gaussian scale model and through a careful selection of a finite collection of least favorable parameters. In addition, optimal rate of convergence for estimating the inverse of a Toeplitz covariance matrix is also established.

**Keywords:** Banding, covariance matrix, minimax lower bound, optimal rate of convergence, spectral norm, tapering, Toeplitz covariance matrix.

**AMS 2000 Subject Classification:** Primary 62H12; secondary 62F12, 62G09

---

<sup>1</sup>Statistics Department, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, The research of Tony Cai was supported in part by NSF FRG Grant DMS-0854973.

<sup>2</sup>Department of Statistics, Yale University, New Haven, CT 06511. The research of Zhao Ren and Harrison Zhou was supported in part by NSF Career Award DMS-0645676 and NSF FRG Grant DMS-0854975.

# 1 Introduction

Estimation of a Toeplitz covariance matrix and its inverse arises naturally in the analysis of stationary time series which are used in a wide range of applications in many fields including engineering, economics, and biology. For example, stationary Gaussian processes is one of the most fundamental models in statistical signal processing and Toeplitz covariance matrices are used for radar imaging, target detection, speech recognition, and communications systems. See, e.g., Snyder, O'Sullivan and Miller (1989), Fuhrmann (1991), Roberts and Ephraim (2000), and Christensen (2007). Toeplitz matrices are also used to model the correlation of cyclostationary processes in periodic time series (Dzhaparidze (1986), Chakraborty (1998), Brockwell and Davis (1991)).

In the classical low dimensional setting, many methods including the maximum likelihood estimator using the EM algorithm have been developed for estimating Toeplitz covariance matrices. However, in the high-dimensional setting, which is becoming increasingly common in many contemporary applications, the standard estimators do not provide satisfactory performance and regularization is needed. In recent papers, Wu and Pourahmadi (2003) has introduced and studied banding estimators for autocovariance matrix of a stationary process which is Toeplitz, and McMurphy and Politis (2010) extended their results to tapering estimators.

The problem of optimal estimation of large covariance matrices has drawn considerable recent attention. In the present paper, we consider estimation of large Toeplitz covariance matrix and its inverse under the matrix spectral norm in the high dimensional setting. The goal is to gain fundamental understanding of the problem by constructing rate-optimal estimators and establishing the optimal rate of convergence. To be more specific, suppose we observe independent and identically distributed (i.i.d.)  $p$ -variate random variables  $\mathbf{X}_1, \dots, \mathbf{X}_n$  with covariance matrix  $\Sigma_{p \times p}$  where each  $\mathbf{X}_i$  is a stationary process so that  $\Sigma_{p \times p}$  has a Toeplitz structure,

$$\Sigma_{p \times p} = \begin{pmatrix} \sigma_0 & \sigma_1 & \sigma_2 & \cdots & \sigma_{p-2} & \sigma_{p-1} \\ \sigma_1 & \sigma_0 & \sigma_1 & & & \sigma_{p-2} \\ \sigma_2 & \sigma_1 & \sigma_0 & & & \vdots \\ \vdots & & & \ddots & & \vdots \\ \sigma_{p-2} & & & & \sigma_0 & \sigma_1 \\ \sigma_{p-1} & \sigma_{p-2} & \cdots & \cdots & \sigma_1 & \sigma_0 \end{pmatrix}. \quad (1)$$

The goal is to estimate the unknown Toeplitz matrix  $\Sigma_{p \times p}$  and its inverse under the

spectral norm based on the sample  $\{\mathbf{X}_i : i = 1, \dots, n\}$ . We assume both  $n$  and  $p$  are growing. However, unlike many other covariance matrix estimation problems, the results also hold for a fixed sample size  $n$ . For example,  $n$  can be taken to be 1 as is common in time series analysis. For a matrix  $A$  its spectral norm is defined as  $\|A\| = \sup_{\|x\|_2=1} \|Ax\|_2$ . The minimax risk of estimating  $\Sigma$  over a given collection  $\mathcal{F}$  of Toeplitz covariance matrices under the spectral norm  $\|\cdot\|$  is defined as

$$R(\mathcal{F}) = \inf_{\hat{\Sigma}} \sup_{\Sigma \in \mathcal{F}} \mathbb{E} \|\hat{\Sigma} - \Sigma\|^2.$$

In the present paper, we establish the optimal rates of convergence of  $R(\mathcal{F})$  over two commonly used parameter spaces and introduce a rate-optimal tapering estimator.

It is clear that the Toeplitz covariance matrix  $\Sigma_{p \times p}$  is uniquely determined by the sequence of covariances  $(\sigma_m) \equiv (\sigma_0, \sigma_1, \dots, \sigma_{p-1}, \dots)$ . A natural parameter space to consider is the following collection defined in terms of the rate of decay of the covariance sequence  $(\sigma_m)$ ,

$$\mathcal{G}_\beta(M) = \left\{ \Sigma_{p \times p} : |\sigma_m| \leq M(m+1)^{-\beta-1}, \Sigma \succ 0 \right\} \quad (2)$$

where  $0 < \beta$ ,  $M < \infty$ , and  $\Sigma \succ 0$  denotes that  $\Sigma$  is positive-semidefinite. It is also well known that the Toeplitz covariance matrix  $\Sigma$  is closely connected to the spectral density of the stationary process  $\mathbf{X}_1$  given by

$$f(x) = \frac{1}{2\pi} \left[ \sigma_0 + 2 \sum_{m=1}^{\infty} \sigma_m \cos(mx) \right], \quad x \in [-\pi, \pi],$$

which is a real-valued and even function on  $[-\pi, \pi]$ . Another natural parameter space to consider is a set defined in terms of the smoothness of the spectral density  $f$ . The parameter space  $\mathcal{F}_\beta(M_0, M)$ , defined in Section 2, contains Toeplitz covariance matrices whose corresponding spectral density functions are of Hölder smoothness  $\beta$ .

Our analysis establishes the minimax rates of convergence for estimating the Toeplitz covariance matrices over the parameter spaces  $\mathcal{G}_\beta(M)$  and  $\mathcal{F}_\beta(M_0, M)$ . We first introduce the tapering and banding estimators and study in detail their properties under the matrix spectral norm. The optimal tapering estimator is constructed and its rate of convergence is derived. Somewhat surprisingly, our results show that the banding estimators and tapering estimators are fundamentally different in the context of estimating Toeplitz covariance matrices over a range of parameter spaces  $\mathcal{F}_\beta(M_0, M)$ , in the sense that the best banding estimator cannot achieve the same rate of convergence as the one attained by the optimal tapering estimator because of a large bias. In other words, banding is strictly sub-optimal and in particular is not as good as tapering for estimating Toeplitz covariance

matrices over a range of  $\mathcal{F}_\beta(M_0, M)$ . However for estimation over the parameter spaces  $\mathcal{G}_\beta(M)$ , with the same choice of the banding and tapering parameters, the two estimators attain the same rate of convergence. This phenomenon is different from those in the estimation of other types of covariance matrices. In addition, we also establish in this paper the optimal rate of convergence for estimating the inverse of a Toeplitz covariance matrix.

The problem of estimating Toeplitz covariance matrices and its inverse exhibits interesting new features different from those in other related covariance matrix estimation problems. This is particularly true for establishing minimax lower bounds. In this paper, the lower bound is obtained through a novel construction of a more informative experiment which is shown to be exactly equivalent to a Gaussian scale model. A minimax lower bound for the more informative model, which immediately provides a lower bound for the original problem, is derived by carefully constructing a collection of least favorable spectral densities and by applying Fano's Lemma. This two-step technique is quite different from those used to establish the optimal rate of convergence in other covariance matrix estimation problems. See, e.g., Cai, Zhang and Zhou (2010) and Cai and Zhou (2010).

By combining the minimax lower and upper bounds developed in later sections, the main results on the optimal rate of convergence for estimating a Toeplitz covariance matrix can be summarized in the following theorem. Here for two sequences of positive numbers  $a_n$  and  $b_n$ ,  $a_n \asymp b_n$  means that there exist positive constants  $c$  and  $C$  independent of  $n$  such that  $c \leq a_n/b_n \leq C$ .

**Theorem 1** *The minimax risk of estimating the Toeplitz covariance matrix  $\Sigma_{p \times p}$  over the collections  $\mathcal{G}_\beta(M)$  or  $\mathcal{F}_\beta(M_0, M)$  satisfies*

$$\inf_{\hat{\Sigma}_{p \times p}} \sup_{\mathcal{H}_\beta} \mathbb{E} \left\| \hat{\Sigma}_{p \times p} - \Sigma_{p \times p} \right\|^2 \asymp \left( \frac{\log(np)}{np} \right)^{\frac{2\beta}{2\beta+1}}$$

*under the condition (7), where  $\mathcal{H}_\beta = \mathcal{G}_\beta(M)$ , or  $\mathcal{H}_\beta = \mathcal{F}_\beta(M_0, M)$  defined in (6).*

Harmonic analysis plays a major role in the technical arguments for establishing both the minimax upper and lower bounds.

In addition to the Toeplitz matrices considered in the present paper, estimation of large covariance matrices under other structural assumptions has been actively studied in the recent literature. The most commonly considered assumptions are “sparse”, where only a small number of entries in each row/column are nonzero, and “bandable”, where the entries of the matrix decay as they move away from the diagonal. Many regularization

methods have been proposed and studied under these assumptions. For example, Bickel and Levina (2008a, b) proposed a banding estimator for estimating bandable covariance matrices and a thresholding estimator for sparse covariance matrices and obtained rate of convergence for the two estimators. See also El Karoui (2008) and Lam and Fan (2009). Cai, Zhang and Zhou (2010) established the optimal rates of convergence for estimating bandable covariance matrices and introduced rate-optimal tapering estimators. Cai and Zhou (2010) derived the minimax rate of convergence for estimating sparse covariance matrices under the spectral norm. In particular, a new general lower bound technique was developed. Cai and Liu (2011) introduced an adaptive thresholding procedure for estimating sparse covariance matrices that automatically adjusts to the variability of individual entries. Estimation of sparse inverse covariance matrices has also drawn considerable attention due to its close connections to Gaussian graphical model selection. See Ravikumar, Wainwright, Raskutti and Yu (2008), Yuan (2010), and Cai, Liu and Luo (2010). The optimal rate of convergence for estimating sparse inverse covariance matrices was established in Cai, Liu and Zhou (2010).

The rest of the paper is organized as follows. In Section 2, tapering and banding estimators are introduced and studied. In particular, a minimax upper bound for estimating Toeplitz covariance matrices under the spectral norm is obtained. Section 3 establishes a minimax lower bound which matches in terms of the rate of convergence the minimax upper bound derived in Section 2. The upper and lower bounds together yield the optimal rate of convergence. Section 4 considers estimation of the inverse of a Toeplitz covariance matrix and establishes the optimal rate of convergence for estimating the inverse under the spectral norm. Section 5 discusses connections and differences of our work with other related problems. The proofs are given in Sections 6 and 7.

## 2 Methodology and Minimax Upper Bound under the Spectral Norm

In this section we introduce tapering and banding procedures for estimating the Toeplitz covariance matrix  $\Sigma_{p \times p}$  based on a random sample of  $p$ -variate Gaussian observations  $\mathbf{X}_1, \dots, \mathbf{X}_n \stackrel{iid}{\sim} N(\mu, \Sigma_{p \times p})$ . The properties of the tapering and banding estimators under the spectral norm are then studied and used to establish the minimax upper bounds.

Given a random sample  $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$  from a Gaussian distribution with a Toeplitz

covariance matrix  $\Sigma_{p \times p}$ , the sample covariance matrix is

$$\Sigma_{p \times p}^* = (\sigma_{st}^*)_{1 \leq s, t \leq p} = \frac{1}{n-1} \sum_{l=1}^n (\mathbf{X}_l - \bar{\mathbf{X}}) (\mathbf{X}_l - \bar{\mathbf{X}})^T \quad (3)$$

where  $\bar{\mathbf{X}} = \frac{1}{n} \sum_{l=1}^n \mathbf{X}_l$  is the sample mean. Note that  $\Sigma^*$  is translation invariant, thus we shall assume  $\mathbb{E}\mathbf{X}_l = 0$  hereafter. When the covariance matrix  $\Sigma_{p \times p}$  is Toeplitz, an immediate improvement of the sample covariance estimator is to average the entries in the diagonals of  $\Sigma_{p \times p}^*$ . For  $0 \leq m \leq p-1$ , set

$$\tilde{\sigma}_m = \frac{1}{p-m} \sum_{s-t=m} \sigma_{st}^* \quad (4)$$

and define the Toeplitz matrix  $\tilde{\Sigma}$  by  $\tilde{\Sigma} = (\tilde{\sigma}_{st})_{1 \leq s, t \leq p}$  with  $\tilde{\sigma}_{st} = \tilde{\sigma}_{|s-t|}$ . Then  $\tilde{\Sigma}$  is an unbiased estimator of  $\Sigma$ .

We shall construct tapering estimators of the Toeplitz covariance matrix  $\Sigma_{p \times p}$  based on the unbiased estimator  $\tilde{\Sigma}$  as follows. For a given even positive integer  $k \leq p/2$ , let  $\omega = (\omega_m)_{0 \leq m \leq p-1}$  be a weight sequence with the  $\omega_m$  given by

$$\omega_m = \begin{cases} 1, & \text{when } m \leq k/2 \\ 2 - \frac{2m}{k}, & \text{when } k/2 < m \leq k \\ 0, & \text{Otherwise} \end{cases}.$$

Define the tapering estimator  $\hat{\Sigma}_k$  of the Toeplitz matrix  $\Sigma$  by  $\hat{\Sigma}_k = (\hat{\sigma}_{st})$  where

$$\hat{\sigma}_{st} = \hat{\sigma}_{|s-t|} = \omega_{|s-t|} \tilde{\sigma}_{|s-t|}.$$

For the tapering estimator it is easy to see

$$\mathbb{E}\hat{\sigma}_m = \omega_m \sigma_m.$$

Similarly, for a given integer  $0 \leq k \leq p-1$ , a banding estimator  $\hat{\Sigma}_k^B$  can be defined as  $\hat{\Sigma}_k^B = (\hat{\sigma}_{st}^B)$  with

$$\hat{\sigma}_{st}^B = \hat{\sigma}_{|s-t|}^B = \tilde{\sigma}_{|s-t|} \cdot I(|s-t| \leq k). \quad (5)$$

It is clear that the tapering estimator  $\hat{\Sigma}_k$  is different from the banding estimator  $\hat{\Sigma}_k^B$ , which is an unbiased estimator of  $\Sigma_k^B = (\sigma_{st}^B)_{1 \leq s, t \leq p}$  with  $\sigma_{st}^B = \sigma_{|s-t|} I(|s-t| \leq k)$ . Note that both tapering and banding estimators have been used for other covariance estimation problems and the two estimators share similar properties. See, e.g., Bickel and Levina (2008a) and Cai, Zhang and Zhou (2010).



As mentioned in the introduction, the Toeplitz covariance matrix is closely connected to the spectral density of the stationary process. In addition to the parameter space  $\mathcal{G}_\beta(M)$  defined in (2) in terms of the rate of decay of the covariance sequence  $(\sigma_m)$ , another natural parameter space to consider is defined in terms of the smoothness of the spectral density  $f$ , which is commonly used in the analysis of periodic time series. Let  $\beta = \gamma + \alpha > 0$ , where  $\gamma$  is the largest integer strictly less than  $\beta$ ,  $0 < \alpha \leq 1$ , and  $0 < M_0, M < \infty$ . Define

$$\mathcal{F}_\beta(M_0, M) = \left\{ \Sigma(f) : \|f\|_\infty \leq M_0 \text{ and } \left\| f^{(\gamma)}(\cdot + h) - f^{(\gamma)}(\cdot) \right\|_\infty \leq Mh^\alpha, \Sigma \succ 0 \right\}, \quad (6)$$

where  $\Sigma(f)$  is a  $p \times p$  Toeplitz matrix uniquely determined by Fourier coefficients of  $f$ . The smoothness parameter  $\beta$  of the spectral density  $f$  is closely connected to the rate of decay of the covariances  $\sigma_m$  as  $m$  increases. The optimal rate of convergence for estimating the Toeplitz covariance matrices  $\Sigma_{p \times p}$  over the parameter space  $\mathcal{F}_\beta(M_0, M)$  critically depends on the value of  $\beta$ . For two parameter spaces  $\mathcal{G}_\beta(M)$  and  $\mathcal{F}_\beta(M_0, M)$ , one is not a subclass of the other. Their connections and differences are discussed in Section 5.

We study the performance of both the tapering estimator  $\hat{\Sigma}_k$  and the banding estimator  $\hat{\Sigma}_k^B$  over the two parameter spaces  $\mathcal{F}_\beta(M_0, M)$  and  $\mathcal{G}_\beta(M)$ . The analysis is quite similar for the two parameter spaces, but the asymptotic behaviors of the tapering and banding estimators are more interesting over  $\mathcal{F}_\beta(M_0, M)$  than  $\mathcal{G}_\beta(M)$ . We therefore will mainly focus our analysis on  $\mathcal{F}_\beta(M_0, M)$ . We begin by establishing the following risk bounds for the tapering estimator  $\hat{\Sigma}_k$  under the spectral norm.

**Remark 1** *Throughout the paper we shall assume that*

$$\left( \frac{np}{\log(np)} \right)^{\frac{1}{2\beta+1}} \leq p/2. \quad (7)$$

*The purpose of assumption (7) is to rule out the naive estimator (4). The right hand side  $p/2$  in (7) can of course be replaced by  $cp$  for any positive constant  $c < 1$ .*

To simplify the notation, from now on we shall write  $\Sigma$  for  $\Sigma_{p \times p}$  if the dependence on  $p$  is clearly understood. Throughout the paper we denote by  $C, c, C_1, c_1, C_2, c_2, \dots$  etc. generic constants, not depending on  $n$  or  $p$ , which may vary from place to place. Let  $\lfloor x \rfloor$  denote the largest integer less than or equal  $x$ .

**Theorem 2** *The tapering estimator  $\hat{\Sigma}_k$  of the Toeplitz covariance matrix  $\Sigma$  with  $k \leq p/2$  satisfies*

$$\sup_{\mathcal{F}_\beta(M_0, M)} \mathbb{E} \left\| \hat{\Sigma}_k - \Sigma \right\|^2 \leq C \frac{k \log(np)}{np} + C k^{-2\beta} \quad (8)$$

for some constant  $C > 0$ . Consequently, by setting an optimal choice  $k = k_* \equiv \left\lfloor \left( \frac{np}{\log(np)} \right)^{\frac{1}{2\beta+1}} \right\rfloor$ , we have

$$\sup_{\mathcal{F}_\beta(M_0, M)} \mathbb{E} \left\| \hat{\Sigma}_{k_*} - \Sigma \right\|^2 \leq C_1 \left( \frac{\log(np)}{np} \right)^{\frac{2\beta}{2\beta+1}}. \quad (9)$$

The upper bounds given in Theorem 2 are proved by using the connections between the spectral norm of a Toeplitz matrix  $\Sigma$  and the supnorm of the corresponding spectral density  $f$ . Indeed,

$$\|\Sigma\| \leq 2\pi \|f\|_\infty = \sup_{[-\pi, \pi]} \left| \sigma_0 + 2 \sum_{m=1}^{\infty} \sigma_m \cos mx \right|. \quad (10)$$

See, for example, Chapter 1 of Böttcher and Silbermann (1999). Note that

$$\mathbb{E} \left\| \hat{\Sigma}_k - \Sigma \right\|^2 \leq 2\mathbb{E} \left\| \hat{\Sigma}_k - \mathbb{E}\hat{\Sigma}_k \right\|^2 + 2 \left\| \mathbb{E}\hat{\Sigma}_k - \Sigma \right\|^2.$$

The variance term  $\mathbb{E} \left\| \hat{\Sigma}_k - \mathbb{E}\hat{\Sigma}_k \right\|^2$  and the bias term  $\left\| \mathbb{E}\hat{\Sigma}_k - \Sigma \right\|^2$  can then be bounded from above by the supnorm of the corresponding spectral densities of the Toeplitz matrices  $\hat{\Sigma}_k - \mathbb{E}\hat{\Sigma}_k$  and  $\mathbb{E}\hat{\Sigma}_k - \Sigma$  respectively. For the variance part, we apply a large deviation result for spectral density estimation from Bentus and Rudzkis (1982) and show that

$$\mathbb{E} \left\| \hat{\Sigma}_k - \mathbb{E}\hat{\Sigma}_k \right\|^2 \leq C \frac{k \log(np)}{np}.$$

The upper bound for the bias term  $\left\| \mathbb{E}\hat{\Sigma}_k - \Sigma \right\|^2$  is of order  $k^{-2\beta}$  due to a well known result for the tapering estimators from harmonic analysis. See Zygmund (2002). Set  $k = k_* \equiv \left\lfloor \left( \frac{np}{\log(np)} \right)^{\frac{1}{2\beta+1}} \right\rfloor$ , then the tapering estimator achieves the rate of convergence  $\left( \frac{\log(np)}{np} \right)^{2\beta/(2\beta+1)}$ .

**Remark 2** *The tapering estimator  $\hat{\Sigma}_{k_*}$  in (9) is not guaranteed to be positive semidefinite for a given sample. By using results on circulant matrices, one can construct a new estimator  $\hat{\Sigma}^{New}$  based on  $\hat{\Sigma}_{k_*}$  such that  $\hat{\Sigma}^{New}$  is positive semidefinite, Toeplitz and attains the upper bound in Equation (9). See Section 5 for details.*

We now turn to the performance of the banding estimator. The analysis is similar, but the result is somewhat surprisingly different. It is interesting to note that the best banding estimator is inferior to the optimal tapering estimator for estimating the Toeplitz covariance matrices over  $\mathcal{F}_\beta(M_0, M)$ . Assume that

$$(np \log(np))^{1/(2\beta+1)} = O(p^\kappa) \quad (11)$$

for some  $\kappa < \frac{2}{5}$ . The following theorem is established by extending a major result of Woodroffe and Van Ness (1967) in which a condition similar to (11) was imposed, together with the fact that the banding estimator may have a large bias as shown in Lemma 3. The details are given in Section 6.3.

**Theorem 3** *Under the assumption (11), the banding estimator (5) satisfies*

$$\left(\frac{np}{\log(np)}\right)^{\frac{2\beta}{2\beta+1}} \inf_k \sup_{\mathcal{F}_\beta(M_0, M)} \mathbb{E} \left\| \hat{\Sigma}_k^B - \Sigma \right\|^2 \rightarrow \infty.$$

Let us now consider the parameter space  $\mathcal{G}_\beta(M)$  defined in (2). It can be shown that the tapering estimator attains the same rate of convergence as the one for  $\mathcal{F}_\beta(M_0, M)$ . Furthermore, in contrast to estimation over  $\mathcal{F}_\beta(M_0, M)$ , for estimating  $\Sigma$  over the parameter space  $\mathcal{G}_\beta(M)$  the banding estimator achieves the same rate of convergence as the tapering estimator.

**Theorem 4** *For  $k \leq p/2$ , the tapering estimator  $\hat{\Sigma}_k$  or the banding estimator  $\hat{\Sigma}_k^B$  of the Toeplitz covariance matrix  $\Sigma$  satisfies, for some constant  $C > 0$ ,*

$$\sup_{\Sigma \in \mathcal{G}_\beta(M)} \mathbb{E} \left\| \hat{\Sigma} - \Sigma \right\|^2 \leq C \frac{k \log(np)}{np} + Ck^{-2\beta}$$

where  $\hat{\Sigma} = \hat{\Sigma}_k$  or  $\hat{\Sigma}_k^B$ . Consequently, by setting an optimal choice of  $k = k_* \equiv \left\lfloor \left(\frac{np}{\log(np)}\right)^{\frac{1}{2\beta+1}} \right\rfloor$ , we have

$$\sup_{\Sigma \in \mathcal{G}_\beta(M)} \mathbb{E} \left\| \hat{\Sigma}_{k_*} - \Sigma \right\|^2 \leq C_1 \left(\frac{\log(np)}{np}\right)^{\frac{2\beta}{2\beta+1}}.$$

The parameter spaces  $\mathcal{G}_\beta(M)$  and  $\mathcal{F}_\beta(M_0, M)$  are similar, but they also have subtle differences which lead to distinct risk properties for the banding estimator over the two parameter spaces. For a Toeplitz covariance matrix  $\Sigma \in \mathcal{G}_\beta(M)$ , due to the rate of decay of the sequence of covariances  $(\sigma_i)$ , the bias component of the risk of the banding estimator  $\hat{\Sigma}_k^B$  has the upper bound

$$\sup_{\mathcal{G}_\beta(M)} \mathbb{E} \left\| \hat{\Sigma}_k^B - \Sigma \right\|^2 \leq Ck^{-2\beta},$$

which is the same as that of the tapering estimator  $\hat{\Sigma}_k$  in terms of the rate of convergence. The bias bound above is different from the case of  $\mathcal{F}_\beta(M_0, M)$ , for which as shown in Lemma 3 the banding estimator  $\hat{\Sigma}_k^B$  satisfies

$$\sup_{\mathcal{F}_\beta(M_0, M)} \|\mathbb{E}\hat{\Sigma}_k^B - \Sigma\|^2 \asymp (\log k)^2 k^{-2\beta},$$

whereas the maximum squared bias of the tapering estimator  $\hat{\Sigma}_k$  is of order  $k^{-2\beta}$ . There is no significant difference in the variance behavior between the banding estimator  $\hat{\Sigma}_k^B$  and the tapering estimator  $\hat{\Sigma}_k$ . We shall omit the proof of Theorem 4 for reasons of space.

### 3 Minimax Lower Bound under the Spectral Norm

The problem of optimal estimation of large covariance matrices poses new technical challenges, partly due to the difficulty in obtaining rate-sharp minimax lower bounds. For estimating Toeplitz covariance matrices, it appears difficult to derive a rate-sharp minimax lower bound directly. In this section we shall establish a minimax lower bound for estimating Toeplitz covariance matrices by first constructing a more informative model under which independent random variables are observed, and then deriving a lower bound for the more informative model through an equivalent Gaussian scale model. The minimax lower bound for the more informative model then immediately yields a lower bound for the original problem.

Recall that in the original experiment, we observe an i.i.d. random sample  $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$  from a  $p$ -variate Gaussian distribution with the Toeplitz covariance matrix  $\Sigma$  given as in (1). Now let us consider an “enlarged” experiment in which one observes an i.i.d. random sample  $\{\mathbf{Y}_1, \dots, \mathbf{Y}_n\}$  from a  $(2p-1)$ -variate Gaussian distribution with a circulant covariance matrix  $\check{\Sigma} = \check{\Sigma}_{(2p-1) \times (2p-1)}$  matrix where

$$\check{\Sigma} = \begin{pmatrix} \sigma_0 & \sigma_1 & \sigma_2 & \cdots & \sigma_{p-2} & \sigma_{p-1} & \sigma_{p-1} & \sigma_{p-2} & \cdots & \sigma_2 & \sigma_1 \\ \sigma_1 & \sigma_0 & \sigma_1 & & & \sigma_{p-2} & \sigma_{p-1} & \sigma_{p-1} & & & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma_0 & & & \vdots & \sigma_{p-2} & \sigma_{p-1} & & & \vdots \\ \vdots & & & \ddots & & \vdots & \vdots & & \ddots & & \vdots \\ \sigma_{p-2} & & & & \sigma_0 & \sigma_1 & \sigma_2 & & & \sigma_{p-1} & \sigma_{p-1} \\ \sigma_{p-1} & \sigma_{p-2} & \cdots & \cdots & \sigma_1 & \sigma_0 & \sigma_1 & \sigma_2 & \cdots & \sigma_{p-2} & \sigma_{p-1} \\ & & & & \dots & \dots & \dots & \dots & & & \end{pmatrix}$$

i.e.,

$$(\check{\Sigma})_{st} = \begin{cases} \sigma_{|s-t|} & \text{when } |s-t| \leq p-1 \\ \sigma_{2p-1-|s-t|} & \text{when } p \leq |s-t| \leq 2p-2 \end{cases}.$$

Denote the vector of the first  $p$  coordinates of  $\mathbf{Y}_i$  by  $\mathbf{Y}_i^{(1)}$  and the last  $p-1$  coordinates by  $\mathbf{Y}_i^{(2)}$ . Then  $\mathbf{Y}_i$  can be written as  $\mathbf{Y}_i = (\mathbf{Y}_i^{(1)}, \mathbf{Y}_i^{(2)})$  and  $\mathbf{Y}_i^{(1)}$  has exactly the same distribution as  $\mathbf{X}_i$ . The second experiment with the random sample  $\{\mathbf{Y}_1, \dots, \mathbf{Y}_n\}$  is clearly more informative than the first one with  $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$  because in the second experiment one can always make inference simply based only on  $\{\mathbf{Y}_1^{(1)}, \dots, \mathbf{Y}_n^{(1)}\}$  and ignore  $\{\mathbf{Y}_1^{(2)}, \dots, \mathbf{Y}_n^{(2)}\}$ .

The major advantage of the more informative experiment is that it is easier to analyze. It is important to note that the second experiment in which we observe the random sample  $\{\mathbf{Y}_1, \dots, \mathbf{Y}_n\}$  is exactly equivalent to a Gaussian scale model under which one observes

$$Z_{ij} = S_p(f)^{1/2} \left( \frac{2\pi j}{2p-1} \right) \xi_{ij}, \text{ with } \xi_{ij} \stackrel{iid}{\sim} N(0, 1), \quad (12)$$

for  $|j| \leq p-1$ , and  $i = 1, 2, \dots, n$ . Here

$$S_p(f)(x) = \frac{1}{2\pi} (\sigma_0 + 2 \sum_{m=1}^{p-1} \sigma_m \cos mx) \quad (13)$$

is the partial sum of  $f$  with order  $p$ . This can be seen as follows. Define

$$v_j = \frac{2\pi j}{2p-1} \quad \text{for } |j| \leq p-1.$$

It is well known (see Brockwell and Davis (1991)) that the spectral decomposition of  $\check{\Sigma}$  can be given by

$$\check{\Sigma} = \sum_{|j| \leq p-1} \lambda_j \mathbf{u}_j \mathbf{u}_j'$$

where  $\lambda_j$  are real eigenvalues and  $\mathbf{u}_j$  are real orthonormal eigenvectors. The eigenvalues are

$$\lambda_j = \sum_{|k| \leq p-1} \sigma_k \exp(-iv_j k) = 2\pi S_p(f)(v_j), \quad |j| \leq p-1$$

where  $S_p(f)(x)$  is the  $p$ th order partial sum of  $f$  given in (13). The eigenvectors  $\mathbf{u}_j$  of the circulant matrix  $\check{\Sigma}$  are given by  $\mathbf{u}_0 = (2p-1)^{-1/2}(1, \dots, 1)$  and for  $j = 1, \dots, p-1$ ,

$$\mathbf{u}_j' = \left( \frac{2}{2p-1} \right)^{1/2} (1, \cos(v_j), \cos(2v_j), \dots, \cos((2p-2)v_j)), \quad (14)$$

$$\mathbf{u}_{-j}' = \left( \frac{2}{2p-1} \right)^{1/2} (0, \sin(v_j), \sin(2v_j), \dots, \sin((2p-2)v_j)), \quad (15)$$

which in fact do not depend on the entries  $\sigma_j$  of the matrix  $\check{\Sigma}$ . In particular, the set of eigenvectors do not depend on the set of eigenvalues  $\lambda_j$ . This is the key advantage of working with the circulant matrix  $\check{\Sigma} = \check{\Sigma}_{(2p-1) \times (2p-1)}$  over the Toeplitz matrix  $\Sigma_{p \times p}$ .

Define the  $(2p-1) \times (2p-1)$  orthogonal matrix  $U$  by  $U = (\mathbf{u}_{-(p-1)}, \dots, \mathbf{u}_{(p-1)})$  and set

$$\mathbf{Z}_i = \frac{1}{\sqrt{2\pi}} U' \mathbf{Y}_i, \quad \text{for } i = 1, 2, \dots, n.$$

Note that  $\mathbf{Z}_i$  are independent  $(2p-1)$ -dimensional zero-mean Gaussian variables and each  $\mathbf{Z}_i$  has a diagonal covariance matrix with values  $\lambda_j/2\pi$  along the diagonal. Hence  $\mathbf{Z}_i$  can be equivalently written in the form of the Gaussian scale model given in (12). Notice that the transformation is invertible and independent of the unknown parameter  $f$ , thus the experiment of observing the random sample  $\{\mathbf{Y}_1, \dots, \mathbf{Y}_n\}$  is exactly equivalent to observing  $\{Z_{ij}, |j| \leq p-1, i = 1, \dots, n\}$  under the Gaussian scale model given in (12).

We shall work with the Gaussian scale model (12) to establish a minimax lower bound. It is clear that for any statistical problem an optimal procedure based on a more informative experiment performs at least as well as the best procedure based on a less informative experiment. Hence, for our problem of estimating  $\Sigma$  under the spectral norm, a minimax lower bound for the above more informative model automatically provides a lower bound for the original model. The following lower bound is obtained through this technique.

**Theorem 5** *The minimax risk for estimating the Toeplitz covariance matrix  $\Sigma$  over  $\mathcal{F}_\beta(M_0, M)$  under the spectral norm satisfies*

$$\inf_{\hat{\Sigma}} \sup_{\mathcal{F}_\beta(M_0, M)} \mathbb{E} \left\| \hat{\Sigma} - \Sigma \right\|^2 \geq c \left( \frac{np}{\log(np)} \right)^{-\frac{2\beta}{1+2\beta}}$$

for some constant  $c > 0$ .

After the construction of the more informative model, there are two additional major steps in establishing the minimax lower bound. The first step is to construct a finite collection of least favorable spectral densities to reduce the lower bound problem for estimating  $\Sigma$  over the whole parameter space to the one for estimating the spectral density over this finite parameter space. The second step is to use Fano's Lemma to obtain a lower bound for estimating the spectral density under the Gaussian scale model (12) over the finite parameter space. This lower bound then yields immediately the desired lower bound for the original problem of estimating a Toeplitz covariance matrix under the spectral norm.

Similarly, the same lower bound can be obtained for the parameter space  $\mathcal{G}_\beta(M)$ .

**Theorem 6** *The minimax risk for estimating the covariance matrix  $\Sigma$  over  $\mathcal{G}_\beta(M)$  under the operator norm satisfies*

$$\inf_{\hat{\Sigma}} \sup_{\mathcal{G}_\beta(M)} \mathbb{E} \left\| \hat{\Sigma} - \Sigma \right\|^2 \geq c \left( \frac{np}{\log(np)} \right)^{-\frac{2\beta}{1+2\beta}}.$$

The upper bounds given in Theorems 2 and 4 together with the lower bounds stated in Theorems 5 and 6 show that the minimax risk of estimating the Toeplitz covariance matrix  $\Sigma_{p \times p}$  over the collections  $\mathcal{G}_\beta(M)$  or  $\mathcal{F}_\beta(M_0, M)$  satisfies

$$\inf_{\hat{\Sigma}_{p \times p}} \sup_{\mathcal{H}_\beta} \mathbb{E} \left\| \hat{\Sigma}_{p \times p} - \Sigma_{p \times p} \right\|^2 \asymp \left( \frac{\log(np)}{np} \right)^{\frac{2\beta}{2\beta+1}}, \quad (16)$$

where  $\mathcal{H}_\beta = \mathcal{G}_\beta(M)$  or  $\mathcal{F}_\beta(M_0, M)$ . The results also show that the tapering estimator  $\hat{\Sigma}_k$  with the tapering parameter  $k = \left\lfloor \left( \frac{np}{\log(np)} \right)^{\frac{1}{2\beta+1}} \right\rfloor$  attains the optimal rate of convergence  $\left( \frac{\log(np)}{np} \right)^{\frac{2\beta}{2\beta+1}}$  over both  $\mathcal{G}_\beta(M)$  and  $\mathcal{F}_\beta(M_0, M)$ , while the banding estimator  $\hat{\Sigma}_k^B$  with the same choice of  $k$  is rate optimal over  $\mathcal{G}_\beta(M)$ , but not for  $\mathcal{F}_\beta(M_0, M)$ . These results show subtle differences between tapering and banding estimators and between the two parameter spaces  $\mathcal{G}_\beta(M)$  and  $\mathcal{F}_\beta(M_0, M)$ .

## 4 Estimation of the Inverse Toeplitz Covariance Matrix

As mentioned in the introduction, the inverse  $\Sigma^{-1}$  of the Toeplitz covariance matrix  $\Sigma$  is of significant interest in many applications. The results and analysis given in the last two sections can be extended to establish the optimal rate of convergence for estimating  $\Sigma^{-1}$  under the spectral norm.

For estimating the inverse  $\Sigma_{p \times p}^{-1}$ , we require the minimum value of the spectral density  $f$  to be bounded from below by a positive constant so that the minimum eigenvalue of  $\Sigma_{p \times p}$  is bounded away from zero for all  $p$ . For a given constant  $\delta > 0$ , define

$$L_\delta = \left\{ f : \inf_x f(x) > \frac{\delta}{2\pi} \right\}.$$

Define the parameter spaces

$$\mathcal{P}_\beta = \mathcal{F}_\beta(M_0, M) \cap L_\delta \quad \text{and} \quad \mathcal{Q}_\beta = \mathcal{G}_\beta(M) \cap L_\delta. \quad (17)$$

Recall that for any  $f \in \mathcal{F}_\beta(M_0, M)$ , we have  $\|f\|_\infty \leq M_0$  and for  $f \in \mathcal{G}_\beta(M)$ , we have  $\|f\|_\infty \leq M/(\beta\pi)$ . Note that for every Toeplitz matrix  $\Sigma$

$$\lambda_{\min}(\Sigma) \geq 2\pi \inf_x f(x) \quad (18)$$

where  $\lambda_{\min}(\Sigma)$  denotes the smallest eigenvalue of  $\Sigma$  (cf. Brockwell and Davis (1991), Proposition 4.5.3). Equations (18) and (10) imply

$$\delta < \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq \eta,$$

where  $\eta = 2\pi \max\{M_0, M/(\beta\pi)\}$ .

The following theorem gives the minimax rate of convergence for estimating  $\Sigma^{-1}$ .

**Theorem 7** *The minimax risk of estimating the inverse of the Toeplitz covariance matrix  $\Sigma^{-1}$  over the class  $\mathcal{P}_\beta$  or the class  $\mathcal{Q}_\beta$  defined in (17) satisfies*

$$\inf_{\hat{\Omega}} \sup_{\mathcal{R}_\beta} \mathbb{E} \left\| \hat{\Omega} - \Sigma^{-1} \right\|^2 \asymp \left( \frac{\log(np)}{np} \right)^{\frac{2\beta}{2\beta+1}}$$

where  $\mathcal{R}_\beta = \mathcal{P}_\beta$  or  $\mathcal{Q}_\beta$ .

In fact, the optimal rate of convergence is achieved by the inverse of a slight modification of the tapering estimator  $\hat{\Sigma}_* = \hat{\Sigma}_{k_*}$  with  $k_* = \left\lfloor \left( \frac{np}{\log(np)} \right)^{\frac{1}{2\beta+1}} \right\rfloor$ . Set

$$\tilde{\Sigma} = \begin{cases} \hat{\Sigma}_* & \text{for } \lambda_{\min}(\hat{\Sigma}) \geq \frac{1}{\log(np)} \\ I & \text{otherwise} \end{cases} \quad (19)$$

and let  $\hat{\Omega}_* = \tilde{\Sigma}^{-1}$ . Then  $\hat{\Omega}_*$  is rate-optimal, i.e.,

$$\sup_{\mathcal{R}_\beta} \mathbb{E} \left\| \hat{\Omega}_* - \Sigma^{-1} \right\|^2 \leq C \left( \frac{\log(np)}{np} \right)^{\frac{2\beta}{2\beta+1}}$$

for some constant  $C > 0$ , where in this case  $\mathcal{R}_\beta = \mathcal{P}_\beta$  or  $\mathcal{Q}_\beta$ . The proof can be found in Section 6.6.

## Minimax lower bound for estimating the inverse

It is interesting to note that it is not necessary to have a completely separate lower bound derivation for estimating the inverse  $\Sigma^{-1}$ . The following simple argument yields a minimax lower bound for estimating  $\Sigma^{-1}$  based on the lower bound for  $\Sigma$ , which is already established in Section 3. Let  $\mathcal{R}_\beta = \mathcal{P}_\beta$  or  $\mathcal{Q}_\beta$ . For any estimator  $\hat{\Omega}$  of  $\Sigma^{-1}$ , define

$$\hat{\Sigma}_{proj} = \arg \min_{A \in \mathcal{R}_\beta} \{ \|\hat{\Omega} - A^{-1}\| \}.$$

In other words,  $\hat{\Sigma}_{proj}^{-1}$  is the closest matrix to  $\hat{\Omega}$  such that  $\hat{\Sigma}_{proj}$  is in the parameter space  $\mathcal{R}_\beta$ . The true  $\Sigma$  is in  $\mathcal{R}_\beta$ , so  $\|\hat{\Omega} - \Sigma^{-1}\| \geq \|\hat{\Omega} - \hat{\Sigma}_{proj}^{-1}\|$  and hence

$$2 \left\| \hat{\Omega} - \Sigma^{-1} \right\| \geq \left\| \hat{\Omega} - \Sigma^{-1} \right\| + \left\| \hat{\Omega} - \hat{\Sigma}_{proj}^{-1} \right\| \geq \left\| \Sigma^{-1} - \hat{\Sigma}_{proj}^{-1} \right\|.$$



Also note that

$$\left\| \hat{\Sigma}_{proj} - \Sigma \right\| = \left\| \hat{\Sigma}_{proj}(\Sigma^{-1} - \hat{\Sigma}_{proj}^{-1})\Sigma \right\| \leq \left\| \hat{\Sigma}_{proj} \right\| \left\| \Sigma^{-1} - \hat{\Sigma}_{proj}^{-1} \right\| \left\| \Sigma \right\|.$$

Since both  $\hat{\Sigma}_{proj}$  and  $\Sigma$  are in the space  $\mathcal{R}_\beta$ , their spectral norms are bounded from above by a constant  $\eta$  as commented earlier, we conclude that

$$\left\| \Sigma^{-1} - \hat{\Sigma}_{proj}^{-1} \right\| \geq \eta^{-2} \cdot \left\| \hat{\Sigma}_{proj} - \Sigma \right\|.$$

Therefore the minimax risk for estimating  $\Sigma^{-1}$  can be bounded from below as

$$\begin{aligned} \inf_{\hat{\Omega}} \sup_{\mathcal{R}_\beta} \mathbb{E} \left\| \hat{\Omega} - \Sigma^{-1} \right\|^2 &\geq \frac{1}{4\eta^4} \inf_{\hat{\Omega}} \sup_{\mathcal{R}_\beta} \mathbb{E} \left\| \hat{\Sigma}_{proj} - \Sigma \right\|^2 \geq \frac{1}{4\eta^4} \inf_{\hat{\Sigma}} \sup_{\mathcal{R}_\beta} \mathbb{E} \left\| \hat{\Sigma} - \Sigma \right\|^2 \\ &\geq c \left( \frac{np}{\log(np)} \right)^{-\frac{2\beta}{2\beta+1}} \end{aligned}$$

for some constant  $c > 0$ .

Note that the above simple argument can also be applied to some other covariance matrix estimation problems such as that in Cai, Zhang and Zhou (2010) and Cai and Zhou (2011) to more conveniently establish a minimax lower bound for estimating the inverse covariance matrices.

## 5 Discussions

This paper introduces a rate optimal tapering estimator and establishes the minimax rate of convergence for estimating Toeplitz covariance matrices over the parameter spaces  $\mathcal{F}_\beta(M_0, M)$  and  $\mathcal{G}_\beta(M)$  under the spectral norm. The results also show interesting differences between the tapering and banding estimators for estimation over  $\mathcal{F}_\beta(M_0, M)$ . A key step in the lower bound argument is the construction of a more informative model for which the minimax lower bound is easier to obtain. The more informative model is shown to be equivalent to a Gaussian scale model and the lower bound for this model is established by carefully constructing a collection of least favorable spectral densities and by applying Fano's Lemma. Harmonic analysis plays a major role in the technical arguments for establishing both the minimax upper and lower bounds.

The problem of estimating Toeplitz covariance matrices is quite distinct from other covariance matrix estimation problems such as those of estimating bandable or sparse covariance matrices. In those problems technical analyses rely much more heavily on random matrix theory and the lower bound techniques are significantly different from what

is used here. See Cai, Zhang and Zhou (2010) and Cai and Zhou (2010). For example, here a major step in the lower bound argument is the construction of a more informative experiment. To the best of our knowledge, this is not needed in other covariance matrix estimation problems.

As mentioned in Section 2, the tapering estimator  $\hat{\Sigma}_{k_*}$  in (9) is not guaranteed to be positive semidefinite for a given realization. Through a circulant matrix, a new estimator  $\hat{\Sigma}^{New}$  can be constructed such that it is positive semidefinite, Toeplitz and attains the upper bound in (9). The construction is as follows. Recall that for the tapering estimator  $\hat{\Sigma}_{k_*}$  with  $k_* = \left\lfloor \left( \frac{np}{\log(np)} \right)^{\frac{1}{2\beta+1}} \right\rfloor$ , the corresponding spectral density is  $\hat{f}_{k_*}(x) = \frac{1}{2\pi} \left( \hat{\sigma}_0 + 2 \sum_{m=1}^{k_*} \hat{\sigma}_m \cos(mx) \right)$ . Define

$$\hat{f}^{New}(x) = \begin{cases} \hat{f}_{k_*}(x), & \text{if } \hat{f}_{k_*}(x) \geq 0 \\ 0, & \text{otherwise} \end{cases}.$$

Let

$$\hat{\Sigma}_{(2p-1) \times (2p-1)}^{New} = 2\pi \sum_{|j| \leq p-1} \hat{f}^{New}(v_j) \mathbf{u}_j \mathbf{u}_j' \quad (20)$$

where  $v_j = (2\pi j)/(2p-1)$  and  $\mathbf{u}_j$  are defined in Equations (14) and (15). Now define a new estimator  $\hat{\Sigma}_{p \times p}^{New}$  by selecting the first  $p$  rows and  $p$  columns of  $\hat{\Sigma}_{(2p-1) \times (2p-1)}^{New}$ . It is clear that  $\hat{\Sigma}_{p \times p}^{New}$  is a Toeplitz matrix. Since  $\hat{f}^{New}(x) \geq 0$ ,  $\hat{\Sigma}_{(2p-1) \times (2p-1)}^{New}$  is non-negative which implies  $\hat{\Sigma}_{p \times p}^{New}$  is non-negative too. The following proposition shows that it attains the optimal rate of convergence  $\left( \frac{np}{\log(np)} \right)^{-2\beta/(2\beta+1)}$ .

**Proposition 1** *The estimator  $\hat{\Sigma}_{p \times p}^{New}$  satisfies*

$$\sup_{\mathcal{H}_\beta} \mathbb{E} \left\| \hat{\Sigma}_{p \times p}^{New} - \Sigma \right\|^2 \leq C \left( \frac{\log(np)}{np} \right)^{2\beta/(2\beta+1)},$$

where  $\mathcal{H}_\beta = \mathcal{G}_\beta(M)$  or  $\mathcal{F}_\beta(M_0, M)$ .

The parameter spaces  $\mathcal{F}_\beta(M_0, M_1)$  and  $\mathcal{G}_\beta(M)$  are similar, but also have subtle differences, which lead to different risk properties for the banding estimators over these two parameters spaces. For any  $M > 0$  and noninteger  $\beta > 0$ , it can be shown that there exists some constants  $M_0$  and  $M_1$  depending on  $M$  such that  $\mathcal{G}_\beta(M) \subset \mathcal{F}_\beta(M_0, M_1)$ . However in general this is not true for integer  $\beta$ . See, for example, Zygmund (2002). Conversely, it is easy to see for any  $\Sigma \in \mathcal{F}_\beta(M_0, M_1)$  we have  $|\sigma_m| \leq M m^{-\beta}$ , where  $M$  is some constant depending only on  $M_0$  and  $M_1$ . Therefore  $\mathcal{F}_\beta(M_0, M_1) \subset \mathcal{G}_{\beta-1}(M)$  for some constant  $M$  depending on  $M_0$  and  $M_1$ .

The problem of estimating a Toeplitz covariance matrix is closely connected to the problem of estimating the spectral densities. For example, an upper bound for the risk of estimating the spectral density  $f$  under the supnorm automatically provides an upper bound for estimating the Toeplitz covariance matrix  $\Sigma$  under the spectral norm through the classical bound

$$\|\hat{\Sigma} - \Sigma\| \leq 2\pi\|\hat{f} - f\|_{\infty}.$$

However, despite their close connections, the two problems are different. For example, it is usually not true that  $\|\hat{\Sigma} - \Sigma\| \geq c\|\hat{f} - f\|_{\infty}$  uniformly over all  $\hat{f}$  and  $f$  for some constant  $c > 0$ . The lower bound argument for the matrix estimation problem is more involved than that for the spectral density estimation problem.

Golubev, Nussbaum and Zhou (2010) studied the asymptotic equivalence between the spectral density estimation and a Gaussian white noise model, which suggests it should be possible to provide an asymptotic equivalence theory for the Toeplitz covariance matrix estimation problem. Observe a sample  $\mathbf{X}_1 = (y(1), \dots, y(p))'$  from a real Gaussian stationary sequence  $y(t)$  with  $Ey(t) = 0$  and autocovariance function  $\sigma_m = Ey(t)y(t+m)$  with the spectral density  $f(x) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \sigma_m \exp(ihx)$ , i.e.,

$$\mathbf{X}_1 \sim N_p(0, \Sigma(f))$$

where  $\Sigma(f)$  is the  $p \times p$  Toeplitz covariance matrix with entries  $(\Sigma)_{j,k} = \sigma_{|k-j|}$ , for  $j, k = 1, \dots, p$ . Let  $\mathcal{F}$  be a set of spectral densities defined by

$$\mathcal{F} = \left\{ f : f \in \mathcal{F}_{\beta}(M_0, M), \text{ and } \inf_x f(x) \geq \epsilon \right\}$$

for  $\beta > 1/2$ , and some positive constants  $\epsilon$ ,  $M$  and  $M_0$ . It was shown in Golubev, Nussbaum and Zhou (2010) that the experiments given by observations

$$\mathbf{X}_1 \sim N_p(0, \Sigma(f))$$

and

$$dZ_x = \log f(x)dx + 2\pi^{1/2}p^{-1/2}dW_x, \quad x \in [-\pi, \pi]$$

with  $f \in \mathcal{F}$  are asymptotically equivalent. This suggests the experiment of observing

$$\mathbf{X}_1, \dots, \mathbf{X}_n \stackrel{i.i.d.}{\sim} N_p(0, \Sigma(f))$$

is asymptotically equivalent to

$$dZ_t = \log f(t)dt + 2\pi^{1/2}(np)^{-1/2}dW_t, \quad t \in [-\pi, \pi]$$

under a certain smoothness assumption. Applications of the asymptotic equivalence theory include sharp asymptotic minimaxity in estimating  $\Sigma$  by expecting that

$$\inf_{\hat{\Sigma}} \sup_{\mathcal{F}} \mathbb{E} \left\| \hat{\Sigma} - \Sigma \right\|^2 = (1 + o(1)) 4\pi^2 \inf_{\hat{f}} \sup_{\mathcal{F}} \mathbb{E} \left\| \hat{f} - f \right\|_{\infty}^2$$

and

$$\inf_{\hat{\Sigma}} \sup_{\mathcal{F}} \mathbb{E} \frac{1}{p} \left\| \hat{\Sigma} - \Sigma \right\|_F^2 = (1 + o(1)) \inf_{\hat{f}} \sup_{\mathcal{F}} \mathbb{E} \left\| \hat{f} - f \right\|_2^2$$

due to the following facts,

$$\|\Sigma_{\infty \times \infty}\| = 2\pi \|f\|_{\infty}$$

and

$$\|f\|_2^2 = \frac{1}{2\pi} \int f^2 = \sum_{m=-\infty}^{\infty} \sigma_m^2 = \sigma_0^2 + 2 \sum_{m=1}^{\infty} \sigma_m^2.$$

It is an interesting and important topic for future research to establish the asymptotic equivalence rigorously.

## 6 Proofs of Main Theorems

In this section, we will first prove the risk upper bounds for the tapering procedures in Sections 6.1 and 6.2, and show that the banding estimator has inferior risk properties in Section 6.3, then establish the minimax lower bounds in Sections 6.4 and 6.5 for the parameter spaces  $\mathcal{F}_{\beta}$  and  $\mathcal{G}_{\beta}$  respectively. In Section 6.6, we prove Theorem 7, which gives minimax risk results for estimating the inverse of a Toeplitz covariance matrix.

### 6.1 Proof of Theorem 2

It follows from the triangle inequality and Equation (10) that

$$\begin{aligned} \left\| \hat{\Sigma}_k - \Sigma \right\|^2 &\leq 2 \left\| \hat{\Sigma}_k - \mathbb{E} \hat{\Sigma}_k \right\|^2 + 2 \left\| \mathbb{E} \hat{\Sigma}_k - \Sigma \right\|^2 \\ &\leq 8\pi^2 \left( \left\| \mathbb{E} \hat{f}_k(x) - \hat{f}_k(x) \right\|_{\infty}^2 + \left\| \mathbb{E} \hat{f}_k(x) - f(x) \right\|_{\infty}^2 \right) \end{aligned}$$

where

$$f(x) = \frac{1}{2\pi} (\sigma_0 + 2 \sum_{m=1}^{\infty} \sigma_m \cos mx), \quad \text{and} \quad \hat{f}_k(x) = \frac{1}{2\pi} (\hat{\sigma}_0 + 2 \sum_{m=1}^k \hat{\sigma}_m \cos mx).$$

We shall establish following upper bounds for the bias and variance separately,

$$\sup_{\mathcal{F}_{\beta}} \left\| \mathbb{E} \hat{f}_k(x) - f(x) \right\|_{\infty}^2 \leq C k^{-2\beta} \quad (21)$$

and

$$\sup_{\mathcal{F}_\beta} \mathbb{E} \left\| \mathbb{E} \hat{f}_k(x) - \hat{f}_k(x) \right\|_\infty^2 \leq C \frac{k \log(np)}{np}. \quad (22)$$

These two bounds together immediately imply Equation (8) of Theorem 2. A trade-off between the bias and variance leads to an optimal choice of  $k = k_* \equiv \left( \frac{np}{\log(np)} \right)^{\frac{1}{2\beta+1}}$ , which yields the rate of convergence  $\left( \frac{\log(np)}{np} \right)^{\frac{2\beta}{2\beta+1}}$  as stated in Equation (9) of Theorem 2.

We now establish Equations (21) and (22). It is relatively easy to derive the upper bound (21) for the bias. Note that

$$\begin{aligned} \mathbb{E} \hat{f}_k(x) &= \frac{1}{2\pi} (\omega_0 \sigma_0 + 2 \sum_{m=1}^k \omega_m \sigma_m \cos mx) \\ &= \frac{1}{2\pi} (\sigma_0 + 2 \sum_{m=1}^{k/2} \sigma_m \cos mx + 2 \sum_{m=k/2+1}^k \left( 2 - \frac{2m}{k} \right) \sigma_m \cos mx). \end{aligned}$$

Since  $\mathbb{E} \hat{f}_k(x)$  is the de la Vallée Poussin mean of  $f$ , we have

$$\left\| \mathbb{E} \hat{f}_k(x) - f(x) \right\|_\infty \leq C \inf_{T \in \text{TriPoly}(k)} \|T - f(x)\|_\infty \quad (23)$$

where  $\text{TriPoly}(k)$  is the collection of all trigonometric polynomial with degree no more than  $k$ , and the right hand side of (23) can be further bounded as

$$\inf_{T \in \text{TriPoly}(k)} \|T - f(x)\|_\infty \leq 3Mk^{-\beta}$$

for  $f \in \mathcal{F}_\beta(M_0, M)$  (cf. Vol 1, Chapter 3.13 and page 117 of Zygmund (2002)). Consequently, we obtain the desired upper bound in Equation (21).

To study the variance part, we need the following large deviation bounds, which is proved in Section 7.1.

**Lemma 1** *For each observation  $\mathbf{X}_l, l = 1, 2, \dots, n$ , the corresponding estimated spectral density  $\hat{f}_k^{(l)}(x)$  has the following property*

$$\mathbb{P} \left\{ \pm \sqrt{\frac{p}{k}} \left( \hat{f}_k^{(l)}(x) - \mathbb{E} \hat{f}_k^{(l)}(x) \right) \geq t \right\} \leq \exp(-c_1 t^2) \quad \text{for} \quad 0 \leq t \leq c_2 \sqrt{\frac{p}{k}}$$

and

$$\mathbb{P} \left\{ \sqrt{\frac{p}{k}} \left| \hat{f}_k^{(l)}(x) - \mathbb{E} \hat{f}_k^{(l)}(x) \right| \geq t \right\} \leq c_3 \exp(-c_4 t)$$

uniformly over all  $x$  and the parameter space  $\mathcal{F}_\beta(M_0, M)$ .

Lemma 1, together with certain continuity property of  $\hat{f}_k(x) = \frac{1}{n} \sum_{l=1}^n \hat{f}_k^{(l)}(x)$ , yields the following desired upper bound for the variance part.

**Lemma 2** *The estimator  $\hat{f}_k$  of spectral density satisfies*

$$\sup_{\mathcal{F}_\beta} \mathbb{E} \left\| \mathbb{E} \hat{f}_k(x) - \hat{f}_k(x) \right\|_\infty^2 \leq C \frac{k \log(np)}{np}.$$

The detailed proofs of Lemmas 1 and 2 are given in Section 7. ■

## 6.2 Proof of Proposition 1

Similar to the definitions of  $\hat{\Sigma}_{(2p-1) \times (2p-1)}^{New}$  and  $\hat{\Sigma}_{p \times p}^{New}$  in Section 5, we define  $\Sigma_{(2p-1) \times (2p-1)}^{Taper}$  by

$$\Sigma_{(2p-1) \times (2p-1)}^{Taper} = 2\pi \sum_{|j| \leq p-1} \mathbb{E} \hat{f}_{k_*}(v_j) \mathbf{u}_j \mathbf{u}_j',$$

where  $v_j = (2\pi j)/(2p-1)$  and  $\mathbf{u}_j$  are defined in Equations (14) and (15), and define a new matrix  $\Sigma_{p \times p}^{Taper}$  by selecting the first  $p$  rows and columns of  $\Sigma_{(2p-1) \times (2p-1)}^{Taper}$ . Note that  $\Sigma_{p \times p}^{Taper} = \mathbb{E} \hat{\Sigma}_{k_*}$ , where  $k_* \equiv \left( \frac{np}{\log(np)} \right)^{\frac{1}{2\beta+1}}$ , then

$$\left\| \Sigma_{p \times p}^{Taper} - \Sigma_{p \times p} \right\|^2 \leq (2\pi)^2 \left\| f - \mathbb{E} \hat{f}_{k_*} \right\|_\infty^2 \leq C k_*^{-2\beta} = C \left( \frac{\log(np)}{np} \right)^{\frac{2\beta}{2\beta+1}}$$

from Theorem 2. By the triangle inequality, we have

$$\left\| \hat{\Sigma}_{p \times p}^{New} - \Sigma_{p \times p} \right\|^2 \leq 2 \left\| \hat{\Sigma}_{p \times p}^{New} - \Sigma_{p \times p}^{Taper} \right\|^2 + 2 \left\| \Sigma_{p \times p}^{Taper} - \Sigma_{p \times p} \right\|^2,$$

thus it is enough to show that

$$\mathbb{E} \left\| \hat{\Sigma}_{p \times p}^{New} - \Sigma_{p \times p}^{Taper} \right\|^2 \leq C \left( \frac{\log(np)}{np} \right)^{2\beta/(2\beta+1)} \quad (24)$$

to establish Proposition 1.

Now we establish Equation (24). Note that

$$\begin{aligned} \left\| \hat{\Sigma}_{p \times p}^{New} - \Sigma_{p \times p}^{Taper} \right\|^2 &\leq \left\| \hat{\Sigma}_{(2p-1) \times (2p-1)}^{New} - \Sigma_{(2p-1) \times (2p-1)}^{Taper} \right\|^2 \\ &= (2\pi)^2 \left( \max_j \left| \hat{f}^{New}(v_j) - \mathbb{E} \hat{f}_{k_*}(v_j) \right| \right)^2 \leq (2\pi)^2 \left\| \hat{f}^{New} - \mathbb{E} \hat{f}_{k_*} \right\|_\infty^2 \end{aligned} \quad (25)$$

By the triangle inequality, we can write

$$\left\| \hat{f}^{New} - \mathbb{E} \hat{f}_{k_*} \right\|_\infty^2 \leq 2 \left\| \hat{f}^{New} - f \right\|_\infty^2 + 2 \left\| f - \mathbb{E} \hat{f}_{k_*} \right\|_\infty^2.$$

Since  $f$  is non-negative and  $\hat{f}^{New}$  is the positive part of  $\hat{f}_{k*}$ , it is easy to see that

$$\left\| \hat{f}^{New} - f \right\|_{\infty}^2 \leq \left\| \hat{f}_{k*} - f \right\|_{\infty}^2,$$

then we have

$$\mathbb{E} \left\| \hat{f}^{New} - \mathbb{E} \hat{f}_{k*} \right\|_{\infty}^2 \leq 2\mathbb{E} \left\| \hat{f}_{k*} - f \right\|_{\infty}^2 + 2\mathbb{E} \left\| f - \mathbb{E} \hat{f}_{k*} \right\|_{\infty}^2 \leq 2\mathbb{E} \left\| \hat{f}_{k*} - f \right\|_{\infty}^2 + Ck_*^{-2\beta}$$

which, together with Equation (25), immediately implies

$$\begin{aligned} \mathbb{E} \left\| \hat{\Sigma}_{p \times p}^{New} - \Sigma_{p \times p}^{Taper} \right\|^2 &\leq (2\pi)^2 \mathbb{E} \left\| \hat{f}^{New} - \mathbb{E} \hat{f}_{k*} \right\|_{\infty}^2 \\ &\leq 8\pi^2 \mathbb{E} \left\| \hat{f}_{k*} - f \right\|_{\infty}^2 + Ck_*^{-2\beta} \leq C \left( \frac{\log(np)}{np} \right)^{2\beta/(2\beta+1)} \end{aligned}$$

where the last inequality follows from Theorem 2.  $\blacksquare$

### 6.3 Proof of Theorem 3

This theorem is a consequence of the following three auxiliary lemmas. The proofs of the first two lemmas can be found in the Appendix. We omit the proof of the third lemma, since it is similar to the tapering case which was shown in Section 6.1. A key step in the proof of Lemma 3 is to follow an example in page 315 of Zygmund (2002) by explicitly constructing a covariance matrix  $\Sigma$ , or equivalently the corresponding spectral density, for which the bias of the banding estimator  $\hat{\Sigma}_k^B$  is much larger than  $k^{-2\beta}$ . Lemma 4 is an extension of a major result in Woodroffe and Van Ness (1967).

**Lemma 3** *The bias of the banding estimator  $\hat{\Sigma}_k^B$  in Equation (5) of the Toeplitz covariance matrix  $\Sigma$  with  $k \leq \frac{p}{2}$  satisfies*

$$\sup_{\mathcal{F}_{\beta}(M_0, M)} \left\| \mathbb{E} \hat{\Sigma}_k^B - \Sigma \right\|^2 \geq Ck^{-2\beta} (\log k)^2$$

for some constant  $C > 0$ .

**Lemma 4** *Let  $\Sigma = I_{p \times p}$ , the identity matrix. The banding estimator  $\hat{\Sigma}_k^B$  with  $k = O(p^{\kappa})$  for some  $\kappa < \frac{2}{5}$  and  $k \rightarrow \infty$  as  $p \rightarrow \infty$  satisfies*

$$E \left\| \hat{\Sigma}_k^B - \Sigma \right\|^2 \geq c \frac{k \log k}{np}$$

for some constant  $c > 0$ . Moreover, if  $k \geq p^{\kappa}$ , the banding estimator satisfies

$$E \left\| \hat{\Sigma}_k^B - \Sigma \right\|^2 \geq c \frac{p^{\kappa} \log p}{np}$$

for some constant  $c > 0$ .

**Lemma 5** *The banding estimator  $\hat{\Sigma}_k^B$  defined in (5) satisfies*

$$\sup_{\mathcal{F}_\beta(M_0, M)} \mathbb{E} \left\| \hat{\Sigma}_k^B - \mathbb{E} \hat{\Sigma}_k^B \right\|^2 \leq C \frac{k \log(np)}{np}, \quad (26)$$

for some constant  $C > 0$ .

It suffices to show that for each fixed pair  $(k, p)$  there exists some  $\Sigma$ , or equivalently some  $f \in \mathcal{F}_\beta(M_0, M)$  such that

$$\mathbb{E} \left\| \hat{\Sigma}_k^B - \Sigma \right\|^2 \geq C \left( \frac{\log(np)}{np} \right)^{\frac{2\beta}{2\beta+1}} (\log np)^{\frac{2}{2\beta+1}-\epsilon}$$

for some constants  $C > 0$  and  $\epsilon < \frac{1}{2\beta+1}$ .

Firstly we consider banding  $\hat{\Sigma}_k^B$  estimators with  $k < (np)^{\frac{1}{2\beta+1}} (\log np)^{(\frac{1}{2\beta+1}-\epsilon)}$ . It follows from Lemma 3 and Equation (26) that

$$\begin{aligned} \mathbb{E} \left\| \hat{\Sigma}_k^B - \Sigma \right\|^2 &\geq \left\| \mathbb{E} \hat{\Sigma}_k^B - \Sigma \right\|^2 - \mathbb{E} \left\| \hat{\Sigma}_k^B - \mathbb{E} \hat{\Sigma}_k^B \right\|^2 \\ &\geq C k^{-2\beta} (\log k)^2 - C_1 \frac{k \log(np)}{np}. \end{aligned}$$

Hence, for some  $\epsilon < \frac{1}{2\beta+1}$  and all sufficiently large  $n$  or  $p$ ,

$$\begin{aligned} \mathbb{E} \left\| \hat{\Sigma}_k^B - \Sigma \right\|^2 &\geq C \left( \frac{\log(np)}{np} \right)^{\frac{2\beta}{2\beta+1}} (\log np)^{\frac{2}{2\beta+1}+2\beta\epsilon} - C_1 \left( \frac{\log(np)}{np} \right)^{\frac{2\beta}{2\beta+1}} (\log np)^{\frac{2}{2\beta+1}-\epsilon} \\ &\geq C_2 \left( \frac{\log(np)}{np} \right)^{\frac{2\beta}{2\beta+1}} (\log np)^{\frac{2}{2\beta+1}-\epsilon}. \end{aligned}$$

When  $k \geq (np)^{\frac{1}{2\beta+1}} (\log np)^{(\frac{1}{2\beta+1}-\epsilon)} = O(p^\kappa)$ , let  $\Sigma$  be the identity matrix, then Lemma 4 implies

$$\mathbb{E} \left\| \hat{\Sigma}_k^B - \Sigma \right\|^2 \geq c \frac{k \log k}{np} \geq C_3 \left( \frac{\log(np)}{np} \right)^{\frac{2\beta}{2\beta+1}} (\log np)^{\frac{2}{2\beta+1}-\epsilon}.$$

## 6.4 Proof of Theorem 5

Define  $f_0 = M_0/2$  and  $f_i$  (with period  $2\pi$ ) as follows,

$$f_i = f_0 + \tau \epsilon_{n,p}^\beta \left[ A \left( \frac{x - \epsilon_{n,p}(i - 0.5)}{\epsilon_{n,p}} \right) + A \left( \frac{x + \epsilon_{n,p}(i - 0.5)}{\epsilon_{n,p}} \right) \right], \quad \epsilon_{n,p} = 2\pi/k_* \quad (27)$$

where  $i = 1, 2, \dots, k_*/2$  with  $k_* = \left\lfloor \left( \frac{np}{\log(np)} \right)^{\frac{1}{2\beta+1}} \right\rfloor$ , and  $A(u) = \exp(-\frac{1}{1-4u^2}) 1_{\{|2u|<1\}}$ . It is easy to see that

$$A \in C^\infty(\mathbb{R}) \cap \mathcal{F}_\beta(e^{-1}, 1/2) \quad \text{and} \quad A(x) > 0 \iff x \in (-1/2, 1/2) \quad (28)$$



and  $f_i$  is positive and even, then  $f_i \in \mathcal{F}_\beta(M_0, M)$  by setting  $\tau$  to be a sufficiently small positive constant. Let  $\mathcal{F}_{sub} = \{f_0, f_1, \dots, f_{k_*/2}\}$ .

As we have seen that there is a close connection between autocovariance matrix and spectral density function, now we reduce the lower bound problem for estimating covariance matrix under the spectral norm to the one for estimating spectral density under the supnorm. The careful construction of  $f_i$  in Equation (27) is crucial to establish the following lemma.

**Lemma 6** *There exists some positive constant  $c$  such that*

$$\inf_{\hat{\Sigma}} \sup_{\mathcal{F}_\beta} \mathbb{E} \left\| \hat{\Sigma} - \Sigma \right\|^2 \geq c \inf_{\tilde{f}} \sup_{\mathcal{F}_{sub}} \mathbb{E} \left\| \tilde{f} - f \right\|_\infty^2.$$

It is then enough to show

$$\inf_{\tilde{f}} \sup_{\mathcal{F}_{sub}} \mathbb{E} \left\| \tilde{f} - f \right\|_\infty^2 \geq c \left( \frac{np}{\log(np)} \right)^{-\frac{2\beta}{1+2\beta}} \quad (29)$$

to establish Theorem 5.

We now establish the lower bound for the spectral density estimation in Equation (29). Recall that we have already constructed a more informative model, which is exactly equivalent to a Gaussian scale model where one observes

$$Z_{ij} = S_p(f)^{1/2} \left( \frac{2\pi j}{2p-1} \right) \xi_{ij}, \text{ with } \xi_{ij} \stackrel{iid}{\sim} N(0, 1),$$

for  $|j| \leq p-1$ , and  $i = 1, 2, \dots, n$ . For the above more informative model we will give a lower bound of order  $\left( \frac{np}{\log(np)} \right)^{-\frac{2\beta}{1+2\beta}}$ , which of course is also a lower bound for the original model. It is easy to see that

$$\|f_i - f_j\|_\infty^2 > c_0 \left( \tau \epsilon_{n,p}^\beta \right)^2 \geq c \left( \frac{np}{\log(np)} \right)^{-\frac{2\beta}{1+2\beta}}. \quad (30)$$

In Section 7.4 we prove the following lemma.

**Lemma 7** *Let  $\mathbb{P}_f$  denote the joint distribution of  $(Z_{ij} : i = 1, \dots, n, |j| \leq p-1)$  indexed by function  $f$ . Then*

$$\frac{2}{k_*} \sum_{i=1}^{k/2} K(\mathbb{P}_{f_i}, \mathbb{P}_{f_0}) \leq a \cdot \log k_*, a \in (0, 1/8). \quad (31)$$

By the Fano's lemma (cf. Tsybakov (2009)), Equation (30) and Lemma 7 immediately imply Equation (29), which then yields Theorem 5 together with Lemma 6. ■

## 6.5 Proof of Theorem 6

The proof of Theorem 6 is similar to that of Theorem 5, except that we need to show that the trigonometric coefficients of  $f_i$  belongs to the parameter space  $\mathcal{G}_\beta(M)$ , i.e.,

$$|\sigma_{m,i}| \leq C(\beta) \tau m^{-\beta-1}$$

uniformly for all  $i = 0, 1, \dots, k/2$ , where the constant  $C(\beta)$  only depends on  $\beta$ . Note that

$$\sigma_{m,i} = \int_{(-\pi, \pi]} \tau \epsilon_{n,p}^\beta \left[ A\left(\frac{x - \epsilon_{n,p}(i - 0.5)}{\epsilon_{n,p}}\right) + A\left(\frac{x + \epsilon_{n,p}(i - 0.5)}{\epsilon_{n,p}}\right) \right] \cos(xm) dx.$$

Since the length of the support of  $A(u)$  is 1 and  $A(u) \leq e^{-1}$ , then there exists a set  $I_i$  with measure  $\epsilon_{n,p}$  such that

$$|\sigma_{m,i}| \leq \int_{I_i} \tau \epsilon_{n,p}^\beta \cdot 2e^{-1} dx = \tau \epsilon_{n,p}^\beta \cdot 2e^{-1} \cdot \epsilon_{n,p} = 2\tau e^{-1} \epsilon_{n,p}^{\beta+1}, \quad (32)$$

which implies  $|\sigma_{m,i}| \leq C(\beta) \tau m^{-\beta-1}$  for  $m \leq k$ . For those  $m > k$ , since  $A$  is chosen such that all derivatives are bounded and vanish at  $-1/2$  and  $1/2$ , we apply integration by parts and immediately obtain  $|\sigma_{m,i}| \leq C(\beta) \tau m^{-\beta-1}$ . ■

## 6.6 Proof of Theorem 7

Since we have included the lower bound derivation in Section 4, here we only need to show the upper bound. Note that

$$\|\hat{\Omega}_* - \Sigma^{-1}\|^2 = \|\tilde{\Sigma}^{-1} - \Sigma^{-1}\|^2 = \|\tilde{\Sigma}^{-1}(\Sigma - \tilde{\Sigma})\Sigma^{-1}\|^2 \leq \|\tilde{\Sigma}^{-1}\|^2 \|\tilde{\Sigma} - \Sigma\|^2 \|\Sigma^{-1}\|^2.$$

It follows from the assumption (17) that  $\|\Sigma^{-1}\|^2 \leq C$  for some  $C > 0$ , then we have

$$\mathbb{E} \|\hat{\Omega}_* - \Sigma^{-1}\|^2 \leq C \mathbb{E} \|\tilde{\Sigma}^{-1}\|^2 \|\tilde{\Sigma} - \Sigma\|^2.$$

Let  $R_0 = \mathbb{E} \left\{ \|\tilde{\Sigma}^{-1}\|^2 \|\tilde{\Sigma} - \Sigma\|^2 \cdot I\{\lambda_{\min}(\tilde{\Sigma}) \leq \delta/2\} \right\}$  and write

$$\begin{aligned} \mathbb{E} \|\tilde{\Sigma}^{-1}\|^2 \|\tilde{\Sigma} - \Sigma\|^2 &= \mathbb{E} \left\{ \|\tilde{\Sigma}^{-1}\|^2 \|\tilde{\Sigma} - \Sigma\|^2 \cdot I\{\lambda_{\min}(\tilde{\Sigma}) > \delta/2\} \right\} + R_0 \\ &\leq C_1 \mathbb{E} \|\tilde{\Sigma} - \Sigma\|^2 + R_0 \end{aligned}$$

where  $\lambda_{\min}(\tilde{\Sigma})$  denotes the minimal eigenvalue of  $\tilde{\Sigma}$ . The risk upper bound is then established by showing that

$$R_0 = o\left(\left(\frac{np}{\log(np)}\right)^{-\frac{2\beta}{2\beta+1}}\right) \quad (33)$$

and

$$\mathbb{E} \left\| \tilde{\Sigma} - \Sigma \right\|^2 \leq C \left( \frac{np}{\log(np)} \right)^{-\frac{2\beta}{2\beta+1}}. \quad (34)$$

The following lemma is helpful to establish Equations (33) and (34). Its proof is very similar to that of Lemma 2 and thus omitted.

**Lemma 8** *For any positive constant  $\delta_1$ , the tapering estimator  $\hat{\Sigma}_{k*}$  satisfies*

$$\sup_{\mathcal{R}_\beta} \mathbb{P}(\left\| \hat{\Sigma}_{k*} - \Sigma \right\| > \delta_1) \leq C_D (np)^{-D}$$

for all  $D > 0$ .

It is easy to establish Equation (34). Indeed,

$$\begin{aligned} \mathbb{E} \left\| \tilde{\Sigma} - \Sigma \right\|^2 &\leq \mathbb{E} \left\| \hat{\Sigma}_* - \Sigma \right\|^2 + \left( \mathbb{P}(\lambda_{\min}(\hat{\Sigma}_*) \leq \frac{1}{\log(np)}) \right) \left\| I - \Sigma \right\|^2 \\ &\leq \mathbb{E} \left\| \hat{\Sigma}_* - \Sigma \right\|^2 + C \mathbb{P} \left( \left\| \hat{\Sigma}_* - \Sigma \right\| > \delta - \frac{1}{\log(np)} \right) \\ &\leq C \left( \frac{np}{\log(np)} \right)^{-\frac{2\beta}{2\beta+1}} + o \left( \left( \frac{np}{\log(np)} \right)^{-\frac{2\beta}{2\beta+1}} \right), \end{aligned}$$

where the last inequality follows from Lemma 8.

To show the Equation (33), we apply the Cauchy-Schwarz inequality to  $R_0$  and have

$$\begin{aligned} R_0 &\leq (\mathbb{E} \left\| \tilde{\Sigma}^{-1} \right\|^4 \left\| \tilde{\Sigma} - \Sigma \right\|^4)^{1/2} \cdot (\mathbb{P}(\left\{ \lambda_{\min}(\tilde{\Sigma}) \leq \delta/2 \right\}))^{1/2} \\ &\leq (\log(np))^2 \left( \mathbb{E} \left\| \tilde{\Sigma} - \Sigma \right\|^4 \right)^{1/2} \cdot (\mathbb{P}(\lambda_{\min}(\hat{\Sigma}_*) \leq \delta/2))^{1/2} \end{aligned} \quad (35)$$

where the second inequality follows from the definition of  $\tilde{\Sigma}$  in (19). Since  $\lambda_{\min}(\Sigma) > \delta$  for  $p$  sufficiently large, we have

$$\left\{ \lambda_{\min}(\hat{\Sigma}_*) \leq \delta/2 \right\} \subset \left\{ \left\| \hat{\Sigma}_* - \Sigma \right\| > \delta/2 \right\}$$

then

$$\mathbb{P}(\lambda_{\min}(\hat{\Sigma}_*) \leq \delta/2) \leq \mathbb{P}(\left\| \hat{\Sigma}_* - \Sigma \right\| > \delta/2)$$

which decays to 0 than any polynomial of  $np$  from Lemma 8. It is trivial to see

$$\mathbb{E} \left\| \tilde{\Sigma} - \Sigma \right\|^4 \leq \mathbb{E} \left\| \tilde{\Sigma} - \Sigma \right\|_{Frobenius}^4 \leq Cp^4$$

which, together with Lemma 8 and Equation (35), proves the negligibility of  $R_0$  in Equation (33), thus we complete the proof of Theorem 7. ■

## 7 Proofs of Auxiliary Lemmas

In this section we collect proofs for some auxiliary lemmas.

### 7.1 Proof of Lemma 1

Write

$$\hat{f}_k^{(l)}(x) = \frac{1}{2\pi} \sum_{|m| \leq p} w_{m,p} C_p(m) e^{-ixm},$$

where

$$C_p(m) = \begin{cases} \frac{1}{p} \sum_{s-t=m} \mathbf{X}_l^{(s)} \mathbf{X}_l^{(t)} & \text{for } |m| \leq k-1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$w_{m,p} = \begin{cases} \frac{p}{p-|m|} & \text{for } |m| \leq k/2 \\ \frac{p}{p-|m|} \frac{2(k-|m|)}{k} & \text{for } k/2 < |m| \leq k \\ 0 & \text{otherwise} \end{cases}.$$

Define  $\|W\|_\infty = \frac{1}{2\pi} \sum_{|m| \leq k} w_{m,p}$  and  $\|W\|_2 = \frac{1}{2\pi} \left( \sum_{|m| \leq k} w_{m,p}^2 \right)^{1/2}$ .

The key technical of the proof of Lemma 1 is Theorem 2.1 of Bentkus and Rudzkis (1982), from which we have

$$\mathbf{P} \left\{ \pm(\hat{f}_k^{(l)}(x) - \mathbb{E}\hat{f}_k^{(l)}(x))a_p \geq t \right\} \leq \exp \left\{ -\frac{t^2}{2} G\left(\frac{t}{\Delta}\right) \right\} \quad (36)$$

for all  $x$  and  $t > 0$ , where  $a_p = \frac{\sqrt{p}}{2\sqrt{\pi}\|W\|_2\|f\|_\infty}$ ,  $\Delta = \frac{1}{2\pi} \frac{\|W\|_2\sqrt{p}}{\|W\|_\infty}$ , and

$$G(t) = \begin{cases} 1, & t = 0 \\ 2t^{-2}[t - \log(1+t)], & t > 0 \end{cases}.$$

The proof of Lemma 1 can be completed by studying  $a_p$ ,  $\Delta$  and  $G$  as follows. Note that for any  $f \in \mathcal{F}_\beta(M_0, M)$  we have  $\|f\|_\infty \leq M_0$ , and for  $k \leq p/2$ ,

$$\|W\|_\infty \asymp k \text{ and } \|W\|_2 \asymp k^{1/2} \quad (37)$$

then we have  $a_p \geq c\sqrt{\frac{p}{k}}$  for some  $c > 0$ . Consequently (36) implies that

$$\mathbf{P} \left\{ \pm(\hat{f}_k^{(l)}(x) - \mathbb{E}\hat{f}_k^{(l)}(x))\sqrt{\frac{p}{k}} \geq t \right\} \leq \exp \left\{ -\frac{t^2}{c^2} G\left(\frac{t}{c\Delta}\right) \right\}.$$

Equation (37) implies  $\Delta \asymp \sqrt{\frac{p}{k}}$ , then there exist some constants  $c_1, c_2 > 0$  such that

$$\mathbb{P} \left\{ \pm \sqrt{\frac{p}{k}} \left( \hat{f}_k^{(l)}(x) - \mathbb{E} \hat{f}_k^{(l)}(x) \right) \geq t \right\} \leq \exp(-c_1 t^2), \quad \text{for } 0 \leq t \leq c_2 \sqrt{\frac{p}{k}}. \quad (38)$$

For  $t \geq 1$  it is easy to see that  $tG(\frac{t}{c_2\Delta}) > c_0 > 0$  for some  $c_0 > 0$ . Therefore we conclude that for  $t \geq 1$  there exists some constant  $c_3 > 0$  such that

$$\mathbf{P} \left\{ \pm \left( \hat{f}_k^{(l)}(x) - \mathbb{E} \hat{f}_k^{(l)}(x) \right) \sqrt{\frac{p}{k}} \geq t \right\} \leq \exp(-c_3 t)$$

Clearly, we could choose a large enough constant  $c_4$  to complete our proof, i.e., for all  $t > 0$

$$\mathbf{P} \left\{ \sqrt{\frac{p}{k}} \left| \hat{f}_k^{(l)}(x) - \mathbb{E} \hat{f}_k^{(l)}(x) \right| \geq t \right\} \leq c_4 \exp(-c_3 t) \quad (39)$$

uniformly over  $x$  and the parameter space  $\mathcal{F}_\beta(M_0, M)$ . ■

## 7.2 Proof of Lemma 2

Set  $A$  to be the uniform grids on  $[-\pi, \pi]$  with  $\text{Card}(A) = (np)^5$ , and define

$$G = \left\{ \sup_{x \in A} \left| \hat{f}_k(x) - \mathbb{E} \hat{f}_k(x) \right| \leq b \left( \frac{k}{np} \log(np) \right)^{\frac{1}{2}} \right\}$$

where  $b$  is a positive constant to be specified later. Write

$$\begin{aligned} & \sup_{\mathcal{F}_\beta} \mathbb{E} \left\| \mathbb{E} \hat{f}_k(x) - \hat{f}_k(x) \right\|_\infty^2 \\ & \leq \sup_{\mathcal{F}_\beta} \mathbb{E} \sup_{x \in A} \left| \hat{f}_k(x) - \mathbb{E} \hat{f}_k(x) \right|^2 + \sup_{\mathcal{F}_\beta} \mathbb{E} \sup_{|x-y| \leq 2\pi(np)^{-5}} \left| \hat{f}_k(x) - \mathbb{E} \hat{f}_k(x) - \hat{f}_k(y) + \mathbb{E} \hat{f}_k(y) \right|^2 \\ & = R_1 + R_2 + R_3 \end{aligned} \quad (40)$$

where

$$\begin{aligned} R_1 &= \sup_{\mathcal{F}_\beta} \mathbb{E} \sup_{x \in A} \left| \hat{f}_k(x) - \mathbb{E} \hat{f}_k(x) \right|^2 \{G\}, \quad R_2 = \sup_{\mathcal{F}_\beta} \mathbb{E} \sup_{x \in A} \left| \hat{f}_k(x) - \mathbb{E} \hat{f}_k(x) \right|^2 \{G^c\} \\ R_3 &= \sup_{\mathcal{F}_\beta} \mathbb{E} \sup_{|x-y| \leq \pi(np)^{-5}} \left| \hat{f}_k(x) - \mathbb{E} \hat{f}_k(x) - \hat{f}_k(y) + \mathbb{E} \hat{f}_k(y) \right|^2. \end{aligned}$$

Note that

$$R_1 \leq c \frac{k}{np} \log(np). \quad (41)$$

We will complete the proof of Lemma 2 by showing that

$$R_2 = o \left( \frac{k}{np} \log(np) \right), \quad (42)$$

$$\text{and } R_3 = o \left( \frac{k}{np} \log(np) \right). \quad (43)$$

We first establish Equation (42). Equation (39) of lemma 1 yields

$$\mathbb{E} \left\{ \exp \left( \sqrt{\frac{p}{k}} \left( \hat{f}_k^{(l)}(x) - \mathbb{E} \hat{f}_k^{(l)}(x) \right) t \right) \right\} \leq \exp(c_5 t^2) \quad \text{for} \quad |t| \leq c_6$$

which implies

$$\mathbb{P} \left\{ \sqrt{\frac{np}{k}} \left| \hat{f}_k(x) - \mathbb{E} \hat{f}_k(x) \right| \geq t \right\} \leq 2 \exp(-c_7 t^2) \quad \text{for} \quad t \leq c_8 \sqrt{n}. \quad (44)$$

where  $\hat{f}_k(x) = \frac{1}{n} \sum_{l=1}^n \hat{f}_k^{(l)}(x)$ . See, for example, Chapter 3.4 of Petrov (1975).

When  $\log(np) = o(n)$ , let  $b > \sqrt{(D+5)/c_7}$  for any positive value  $D$ . Equation (44) implies

$$\mathbb{P} \{G^c\} \leq 2(np)^5 \exp(-c_7 b^2 \log(np)) = o((np)^{-D}).$$

If  $\log(p) \geq c_9 n$  for some constant  $c_9 > 0$ , then we have  $\log(np) = o(\frac{p}{k})$  by noting  $k = \left(\frac{np}{\log(np)}\right)^{1/(1+2\beta)}$ . Write  $Z_l = \sqrt{\frac{p}{k}} \left( \hat{f}_k^{(l)}(x) - \mathbb{E} \hat{f}_k^{(l)}(x) \right)$  and its truncation  $Z_l^T = Z_l \mathbb{1}_{\{|Z_l| \leq b_1 \sqrt{\log(np)}\}}$  for some large constant  $b_1$  and each  $l = 1, 2, \dots, n$ . Note that  $Z_l^T - \mathbb{E} Z_l^T$  is subgaussian because of  $\log(np) = o(\frac{p}{k})$  and the tail probability (38), then there is a constant  $c_{10}$  such that

$$\begin{aligned} \mathbb{P} \{G^c\} &\leq (np)^5 \mathbb{P} \left\{ \sqrt{\frac{np}{k}} \left| \hat{f}_k(x) - \mathbb{E} \hat{f}_k(x) \right| \geq b \sqrt{\log(np)} \right\} \\ &\leq (np)^5 \left[ \mathbb{P} \left\{ \frac{1}{\sqrt{n}} \left| \sum_{l=1}^n Z_l^T \right| \geq b \sqrt{\log(np)} \right\} + n \mathbb{P} \left\{ |Z_1| \geq b_1 \sqrt{\log(np)} \right\} \right] \\ &\leq (np)^5 \left[ (np)^{-b^2 c_{10}} + n (np)^{-b_1^2 c_{10}} \right] = o((np)^{-D}). \end{aligned}$$

The last step holds by setting the constants  $b > \sqrt{(D+5)/c_{10}}$  and  $b_1 > \sqrt{(D+5)/c_{10}}$ . Therefore we conclude that  $\mathbb{P} \{G^c\} = o((np)^{-D})$  for any positive value  $D$ .

Moreover,

$$\begin{aligned} \mathbb{E} \sup_x \left| \hat{f}_k(x) - \mathbb{E} \hat{f}_k(x) \right|^4 &= \left( \frac{1}{2\pi} \right)^4 \mathbb{E} \sup_x \left| \hat{\sigma}_0 - \sigma_0 + 2 \sum_{m=1}^k (\hat{\sigma}_m - \omega_m \sigma_m) \cos mx \right|^4 \\ &\leq c \mathbb{E} \left( \sum_{m=0}^k |\hat{\sigma}_m - \omega_m \sigma_m| \right)^4 \leq ck^3 \mathbb{E} \sum_{m=0}^k (\hat{\sigma}_m - \omega_m \sigma_m)^4 \\ &\leq ck^3 \mathbb{E} \sum_{m=0}^k \left( \frac{1}{p-m} \sum_{t-s=m} \sigma_{st}^* - \sigma_m \right)^4 \\ &\leq ck^3 \sum_{m=0}^k \left( \frac{\sum_{t-s=m} \mathbb{E} (\sigma_{st}^* - \sigma_m)^4}{p-m} \right) \\ &= O\left(\frac{k^4}{n^2}\right). \end{aligned}$$

where in the last step we used the normality assumption. The Cauchy-Schwarz inequality then implies

$$\begin{aligned} R_2 &\leq \left[ \mathbb{E} \left( \sup_{x \in A} \left| \hat{f}_k(x) - \mathbb{E} \hat{f}_k(x) \right| \right)^4 \mathbb{P} \{G^c\} \right]^{\frac{1}{2}} \\ &= O \left( \left( \frac{k^4}{n^2} (np)^{-D} \right)^{1/2} \right) = o \left( \frac{k}{np} \log(np) \right) \end{aligned}$$

uniformly over the parameter space  $\mathcal{F}_\beta(M_0, M)$  by letting  $D$  large.

We now establish Equation (43). Note that

$$\begin{aligned} R_3 &\leq \sup_{\mathcal{F}_\beta} \mathbb{E} \sup_{|x-y| \leq 2\pi(np)^{-5}} 2 \left( \sum_{m=1}^{k-1} |\hat{\sigma}_m - \omega_m \sigma_m| |\cos mx - \cos my| \right)^2 \\ &\leq 2k \sup_{\mathcal{F}_\beta} \sum_{m=1}^{k-1} \mathbb{E} \sup_{|x-y| \leq 2\pi(np)^{-5}} |\hat{\sigma}_m - \omega_m \sigma_m|^2 |\cos mx - \cos my|^2. \end{aligned}$$

Since  $\omega_m \leq 1$  and  $|\cos mx - \cos my| \leq m|x-y|$ , we have

$$\begin{aligned} R_3 &\leq \frac{8\pi^2 k}{(np)^{10}} \sup_{\mathcal{F}_\beta} \sum_{m=1}^{k-1} m^2 \mathbb{E} \left( \frac{1}{p-m} \sum_{t-s=m} \sigma_{st}^* - \sigma_m \right)^2 \\ &\leq \frac{ck}{(np)^{10}} \sup_{\mathcal{F}_\beta} \sum_{m=1}^{k-1} m^2 \left[ \frac{1}{p-m} \sum_{t-s=m} \mathbb{E} (\sigma_{st}^* - \sigma_m)^2 \right] \\ &\leq \frac{c_{11}k}{(np)^{10}} \sup_{\mathcal{F}_\beta} \sum_{m=1}^{k-1} m^2 \frac{1}{p-m} \cdot p \\ &\leq \frac{c_{12}k}{(np)^{10}} k^4 = o \left( \frac{k}{np} \log(np) \right). \end{aligned}$$

Equations (40)-(43) all together complete the proof of Lemma 2.  $\blacksquare$

### 7.3 Proof of Lemma 6

Since

$$\sup_{\mathcal{F}_\beta(M_0, M)} \mathbb{E} \left\| \hat{\Sigma} - \Sigma \right\|^2 \geq \sup_{\mathcal{F}_{sub}} \mathbb{E} \left\| \hat{\Sigma} - \Sigma \right\|^2,$$

it is enough to show that

$$\inf_{\hat{\Sigma}} \sup_{\mathcal{F}_{sub}} \mathbb{E} \left\| \hat{\Sigma} - \Sigma \right\|^2 \geq cR(\mathcal{F}_{sub}) \quad (45)$$

to establish Lemma 6, where  $R(\mathcal{F}_{sub}) = \inf_{\tilde{f}} \sup_{\mathcal{F}_{sub}} \mathbb{E} \left\| \tilde{f} - f \right\|_\infty^2$ .

The estimator  $\hat{\Sigma}$  in (45) can be arbitrary, but we show that it is enough to consider estimators of the Toeplitz form in the parameter space  $\mathcal{F}_{sub}$  as follows. For any estimator  $\hat{\Sigma}$ , we define  $\hat{\Sigma}_{sub}$  to be the closest matrix in  $\mathcal{F}_{sub}$  to  $\hat{\Sigma}$  in terms of the spectral norm. For  $\Sigma \in \mathcal{F}_{sub}$  we have

$$2 \left\| \hat{\Sigma} - \Sigma \right\| \geq \left\| \hat{\Sigma} - \Sigma \right\| + \left\| \hat{\Sigma} - \hat{\Sigma}_{sub} \right\| \geq \left\| \hat{\Sigma}_{sub} - \Sigma \right\|$$

which implies

$$\left\| \hat{\Sigma} - \Sigma \right\| \geq \frac{1}{2} \left\| \hat{\Sigma}_{sub} - \Sigma \right\|. \quad (46)$$

Thus a minimax lower bound for estimators of the Toeplitz form in  $\mathcal{F}_{sub}$  provides a lower bound among all possible estimators up to a constant factor 1/2, i.e.,

$$\inf_{\hat{\Sigma}} \sup_{\Sigma \in \mathcal{F}_{sub}} \mathbb{E} \left\| \hat{\Sigma} - \Sigma \right\|^2 \geq \frac{1}{2} \inf_{\hat{\Sigma}_{sub} \in \mathcal{F}_{sub}} \sup_{\Sigma \in \mathcal{F}_{sub}} \mathbb{E} \left\| \hat{\Sigma}_{sub} - \Sigma \right\|^2. \quad (47)$$

To establish Equation (45), it is then sufficient to show that

$$\inf_{\hat{\Sigma}_{sub} \in \mathcal{F}_{sub}} \sup_{\Sigma \in \mathcal{F}_{sub}} \mathbb{E} \left\| \hat{\Sigma}_{sub} - \Sigma \right\|^2 \geq \pi^2 R(\mathcal{F}_{sub}). \quad (48)$$

for  $p$  sufficiently large.

A key tool to establish Equation (48) is the following fact,

$$\left\| \Sigma \right\| \geq \sup_{x \in [-\pi, \pi]} \frac{\langle \Sigma v_x, v_x \rangle}{\langle v_x, v_x \rangle} = 2\pi \sup_{x \in [-\pi, \pi]} |F_p(f)(x)| \quad (49)$$

where  $v_x = (e^{ix}, e^{i2x}, \dots, e^{ipx})$  for any Toeplitz matrix  $\Sigma$  of size  $p \times p$ , and

$$F_p(f)(x) = \frac{1}{2\pi} (\sigma_0 + 2 \sum_{m=1}^{p-1} (1 - \frac{m}{p}) \sigma_m \cos mx).$$

Define

$$\Sigma_{Bk_*} = [\sigma_m \omega_m]_{p \times p}$$

where

$$\omega_m = \begin{cases} 1 & \text{when } m \leq Bk_*/2 \\ 2 - \frac{2m}{Bk_*} & \text{when } Bk_*/2 < m \leq Bk_* \\ 0 & \text{Otherwise} \end{cases}$$

i.e.  $\Sigma_{Bk}$  is a tapering matrix of  $\Sigma$ , and

$$T_{Bk}(f)(x) = \frac{1}{2\pi} (\sigma_0 + 2 \sum_{m=1}^{Bk} \omega_m \sigma_m \cos mx).$$



By the triangle inequality and Equation (49) we have

$$\begin{aligned}
& \left\| \hat{\Sigma}_{sub} - \Sigma \right\| \\
& \geq \left\| \hat{\Sigma}_{sub} - \Sigma_{Bk} \right\| - \left\| \Sigma_{Bk} - \Sigma \right\| \geq 2\pi \left\| F_p(T_{Bk}(f)) - F_p(\hat{f}_{sub}) \right\|_{\infty} - \left\| \Sigma_{Bk} - \Sigma \right\| \\
& \geq 2\pi \left[ \left\| F_p(\hat{f}_{sub}) - f \right\|_{\infty} - \left\| f - T_{Bk}(f) \right\|_{\infty} - \left\| T_{Bk}(f) - F_p(T_{Bk}(f)) \right\|_{\infty} \right] - \left\| \Sigma_{Bk} - \Sigma \right\|, \\
& = 2\pi \left\| F_p(\hat{f}_{sub}) - f \right\|_{\infty} - [2\pi \left\| f - T_{Bk}(f) \right\|_{\infty} + 2\pi \left\| T_{Bk}(f) - F_p(T_{Bk}(f)) \right\|_{\infty} + \left\| \Sigma_{Bk} - \Sigma \right\|]
\end{aligned}$$

From Equation (29) we have seen that

$$\inf_{\tilde{f}} \sup_{\{f_0, f_1, \dots, f_{k/2}\}} (2\pi)^2 \mathbb{E} \left\| F_p(\tilde{f}) - f \right\|_{\infty}^2 \geq 4\pi^2 \inf_{\tilde{f}} \sup_{\mathcal{F}_{sub}} \mathbb{E} \left\| \tilde{f} - f \right\|_{\infty}^2 \geq c_2 \left( \frac{np}{\log(np)} \right)^{-\frac{2\beta}{1+2\beta}},$$

which will be helpful to show that  $2\pi \left\| F_p(\hat{f}_{sub}) - f \right\|_{\infty}$  is the dominating term in Equation (50) as follows. From Equation (21), we have

$$\sup_{\mathcal{F}_{sub}} \left\| \Sigma_{Bk} - \Sigma \right\| \leq CB^{-\beta} k_*^{-\beta} = CB^{-\beta} \left( \frac{np}{\log(np)} \right)^{-\frac{\beta}{1+2\beta}} \quad (51)$$

$$2\pi \sup_{\mathcal{F}_{sub}} \left\| f - T_{Bk}(f) \right\|_{\infty} \leq 2\pi CB^{-\beta} k_*^{-\beta} \left( \frac{np}{\log(np)} \right)^{-\frac{\beta}{1+2\beta}} \quad (52)$$

which can be made to be bounded by  $\varepsilon \sqrt{R(\mathcal{F}_{sub})}$  for any  $\varepsilon > 0$  by setting the constant  $B$  sufficiently large. The term  $\left\| T_{Bk}(f_i) - F_p(T_{Bk}(f_i)) \right\|_{\infty}$  is negligible, since

$$\begin{aligned}
& 2\pi \left\| T_{Bk}(f_i) - F_p(T_{Bk}(f_i)) \right\|_{\infty} \\
& \leq 2 \left\| \sum_{m=1}^{Bk} \omega_m \sigma_{m,i} \frac{m}{p} \cos mx \right\|_{\infty} \\
& \leq 4\tau e^{-1} \epsilon_{n,p}^{\beta+1} \frac{1}{p} \sum_{m=1}^{Bk} m = O \left( \frac{\tau B^2 k^2}{p} \epsilon_{n,p}^{\beta+1} \right) = O \left( \tau \sqrt{R(\mathcal{F}_{sub})} \right)
\end{aligned} \quad (53)$$

where the second inequality is due to the bound  $|\sigma_{m,i}| \leq 2\tau e^{-1} \epsilon_{n,p}^{\beta+1}$  in Equation (32). This value also can be made to be bounded by  $\varepsilon \sqrt{R(\mathcal{F}_{sub})}$  for any  $\varepsilon > 0$  by setting the constant  $\tau$  sufficiently small after setting the constant  $B$ .

Equations (50)-(53) imply (48), which together with Equation (47) yield the proof of Lemma 6. ■

## 7.4 Proof of Lemma 7

Note that  $f_0 = M_0/2$  is a constant function, hence  $S_p(f_0) = f_0$  for all  $p$ . Since

$$K(N(0, \sigma_1), N(0, \sigma_0)) = \frac{1}{2} \left( \frac{\sigma_1}{\sigma_0} - 1 - \log \frac{\sigma_1}{\sigma_0} \right),$$

we have

$$\frac{2}{k} \sum_{i=1}^{k/2} K(\mathbb{P}_{f_i}, \mathbb{P}_{f_0}) = n \frac{2}{k} \sum_{i=1}^{k/2} \sum_{|j| \leq p-1} \frac{1}{2} \left( \frac{S_p(f_i)(t_j)}{f_0(t_j)} - 1 - \log \frac{S_p(f_i)(t_j)}{f_0(t_j)} \right), \quad (54)$$

where  $t_j = \frac{2\pi j}{2p-1}$ . We will show that

$$\left( \frac{S_p(f_i)(t_j)}{f_0(t_j)} - 1 \right) - \log \frac{S_p(f_i)(t_j)}{f_0(t_j)} \leq \frac{4}{M_0^2} (S_p(f_i)(t_j) - f_0(t_j))^2, \text{ for all } i \text{ and } j, \quad (55)$$

and

$$\frac{n}{k} \frac{4}{M_0^2} \sum_{i=1}^{k/2} \sum_{|j| \leq p-1} (S_p(f_i)(t_j) - f_0(t_j))^2 \leq C\tau^2 \log(np), \quad (56)$$

which are crucial to bound (54) and prove Lemma 7.

We first establish Equation (55). Since  $S_p$  is a linear operator, we may write

$$S_p(f_i) = f_0 + \tau \epsilon_{n,p}^\beta \left[ S_p \left[ A \left( \frac{x - \epsilon_{n,p}(i - 0.5)}{\epsilon_{n,p}} \right) \right] + S_p \left[ A \left( \frac{x - \epsilon_{n,p}(i + 0.5)}{\epsilon_{n,p}} \right) \right] \right],$$

then

$$\|S_p(f_i) - f_0\|_\infty \leq 2\tau \epsilon_{n,p}^\beta \left\| S_p \left( A \left( \frac{x}{\epsilon_{n,p}} \right) \right) \right\|_\infty \rightarrow 0.$$

Since  $a - \log(1+a) \leq a^2$  when  $|a| \leq 1/4$ , consequently we have

$$\left( \frac{S_p(f_i)(t_j)}{f_0(t_j)} - 1 \right) - \log \frac{S_p(f_i)(t_j)}{f_0(t_j)} \leq \left( \frac{S_p(f_i)(t_j)}{f_0(t_j)} - 1 \right)^2 = \frac{4}{M_0^2} (S_p(f_i)(t_j) - f_0(t_j))^2.$$

Now we show Equation (56). Recall that  $S_p(f_i)(t_j) - f_0(t_j) = \sum_{m=1}^{p-1} \sigma_{m,i} \cos(t_j m)$ , then we write

$$\frac{n}{k} \frac{4}{M_0^2} \sum_{i=1}^{k/2} \sum_{|j| \leq p-1} (S_p(f_i)(t_j) - f_0(t_j))^2 \leq \frac{n}{k} \frac{4}{M_0^2} \sum_{i=1}^{k/2} \sum_{|j| \leq p-1} \left( \sum_{m=1}^{p-1} \sigma_{m,i} \cos(t_j m) \right)^2.$$

Since

$$\frac{2}{2p-1} \sum_{j=1}^{2p-1} \varphi_m \left( \frac{2\pi j}{2p-1} \right) \varphi_{m'} \left( \frac{2\pi j}{2p-1} \right) = \delta_{mm'}, \quad 1 \leq m, m' \leq p-1,$$

where  $\varphi_m(x) = \cos(xm)$ , the Parseval's identity yields

$$\begin{aligned} \frac{2}{k} \sum_{i=1}^{k/2} K(\mathbb{P}_{f_i}, \mathbb{P}_{f_0}) &\leq \frac{n}{k} \frac{4}{M_0^2} \cdot \frac{2p-1}{2} \sum_{i=1}^{k/2} \sum_{m=1}^{p-1} \sigma_{m,i}^2 \\ &= \frac{n}{k} \frac{4}{M_0^2} \cdot \frac{2p-1}{2} \sum_{i=1}^{k/2} \int_{[-\pi, \pi]} [S_p(f_i - f_0)(x)]^2 dx \end{aligned}$$

which is bounded by

$$\frac{n}{k} \frac{4}{M^2} \frac{2p-1}{2} \sum_{i=1}^{k/2} \int_{[-\pi, \pi]} (f_i(x) - f_0(x))^2 dx \leq Cnp\tau^2 \epsilon_{n,p}^{2\beta+1} = C\tau^2 \log(np).$$

Equations (54)-(56) implies

$$\frac{2}{k} \sum_{i=1}^{k/2} K(\mathbb{P}_{f_i}, \mathbb{P}_{f_0}) \leq C\tau^2 \log(np),$$

which can be bounded by  $a \cdot \log k = \frac{a}{1+2\beta} \log(np) (1 + o(1))$  by choosing  $\tau$  sufficiently small. We then establish Lemma 7. ■

## Acknowledgment

We thank Zongming Ma for a helpful discussion which led to Proposition 1.

## References

- [1] Bentkus, R. Yu. and Rudzkis, R. A. (1982). On the distributions of some statistical estimates of spectral density. *Theory Probab. Appl.* **27**, 795-814.
- [2] Bickel, P. J. and Levina, E. (2008a). Regularized estimation of large covariance matrices. *Ann. Statist.* **36**, 199-227.
- [3] Bickel, P. J. and Levina, E. (2008b). Covariance regularization by thresholding. *Ann. Statist.* **36**, 2577-2604.
- [4] Böttcher, A. and Silbermann, B. (1999) *Introduction to Large Truncated Toeplitz Matrices*. Springer, New York.
- [5] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. 2nd ed. Springer, New York.
- [6] Cai, T. T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.*, to appear.
- [7] Cai, T. T., Liu, W. and Luo, X. (2010). A constrained  $\ell_1$  minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.*, to appear.
- [8] Cai, T. T., Liu, W. and Zhou, H. H. (2010). Optimal estimation of large sparse precision matrices. Manuscript.

- [9] Cai, T. T., Zhang, C.-H. and Zhou, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38**, 2118-2144.
- [10] Cai, T. T. and Zhou, H. H. (2010). Optimal rates of convergence for sparse covariance matrix estimation. Technical report.
- [11] Cai, T. T. and Zhou, H. H. (2011). Minimax estimation of large covariance matrices under  $\ell_1$  norm. *Statistica Sinica*, to appear.
- [12] Chakraborty, M. (1998). An efficient algorithm for solving general periodic Toeplitz system. *IEEE Transactions on Signal Processing* **46**, 784-787.
- [13] Christensen, L. P. B. (2007). An EM-algorithm for band-toeplitz covariance matrix estimation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu, Hawaii, USA.
- [14] Dzhaparidze, K. (1986). *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. Springer, New York.
- [15] El Karoui, N. (2008). Operator norm consistent estimation of large dimensional sparse covariance matrices. *Ann. Statist.* **36**, 2717-2756.
- [16] Fuhrmann, D. R. (1991). Application of Toeplitz covariance estimation to adaptive beamforming and detection. *IEEE Trans. Signal Processing* **39**, 2194-2198.
- [17] Golubev, G. K., Nussbaum, M. and Zhou, H. H. (2010). Asymptotic equivalence of spectral density estimation and Gaussian white noise. *Ann. Statist.* **38**, 181-214.
- [18] Lam, C. and Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrices estimation. *Ann. Statist.* **37**, 4254-4278.
- [19] McMurry, T. L. and Politis, D. N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis* **31**, 471-482.
- [20] Petrov, V. V. (1975). *Sums of Independent Random Variables*, Springer-Verlag, Berlin and New York.
- [21] Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. (2008). High-dimensional covariance estimation by minimizing  $l_1$ -penalized log-determinant divergence. Technical Report, UC Berkeley.

- [22] Roberts, W. J. and Ephraim, Y. (2000). Hidden Markov modeling of speech using Toeplitz covariance matrices. *Speech Communications* **31**, 1-14.
- [23] Snyder, D. L, O’Sullivan, J. A. and Miller, M. I. (1989). The use of maximum-likelihood estimation for forming images of diffuse radar-targets from delay-Doppler data. *IEEE Trans. Inform.Theory* **35**, 536-548.
- [24] Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York.
- [25] Woodroffe, M. B. and W. Van Ness, J. W. (1967). The Maximum Deviation of Sample Spectral Densities. *Ann. Math. Statist.* **38**, 1558-1569.
- [26] Wu, W. B. and Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* **90**, 831-844.
- [27] Yuan, M. (2010). Sparse inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research* **11**, 2261-2286.
- [28] Zygmund, A. (2002). *Trigonometric series. Vol. I, II*. Cambridge Mathematical Library (3rd ed.), Cambridge University Press.

## 8 Appendix

We prove Lemmas 3 and 4 in this appendix.

### 8.1 Proof of Lemma 3

To show for the banding estimator the bias part has lower bound  $k^{-2\beta} (\log k)^2$ , we only need to construct a special spectral density  $f \in \mathcal{F}_\beta(M_0, M)$  for each fixed pair  $(k, p)$  with  $k \leq p/2$  such that  $\left\| \mathbb{E} \hat{\Sigma}_k^B - \Sigma \right\|_\infty \geq ck^{-\beta} (\log k)$  for some positive constant  $c$ . Here we only give a special spectral density for  $0 < \beta < 1$ . The construction is similar to an example in Zygmund (2002) (page 315, example 10). Define the function

$$Q(x, N, n) = 2 \sin Nx \sum_{t=1}^n \frac{\sin(tx)}{t}$$

Clearly

$$Q = \frac{\cos(N-n)x}{n} + \frac{\cos(N-n-1)x}{n-1} + \dots \\ + \frac{\cos(N-1)x}{1} - \frac{\cos(N+1)x}{1} - \dots - \frac{\cos(N+n)x}{n}$$

is a purely cosine polynomial with terms of rank varying from  $N-n$  to  $N+n$ . On one hand the polynomial  $Q$  is uniformly bounded in  $x, N, n$ , say  $\|Q\|_\infty \leq A$ . On the other hand, at  $x=0$  the sum of the first  $n$  terms of  $Q(x, N, n)$  is  $1/n + \dots + 1/2 + 1 > \log n$ . For each pair  $(k, p)$  with  $k \leq p/2$ , let us define

$$f(x) = 2A + 4^{-t\beta} Q(x, k, 4^t) \quad (57)$$

with  $k \in [4^t, 4^{t+1})$ . Clearly,  $A \leq f(x) \leq 3A$ , therefore it's indeed a spectral density since the Toeplitz matrix  $\Sigma_{p \times p}$  corresponding to  $f(x)$  is positive definite for any  $p$ .

It's not hard to check that for each  $M_0$  and  $M$ , we may pick a constant  $C > 0$  such that  $Cf(x) \in \mathcal{F}_\beta(M_0, M)$  uniformly for all pairs  $(k, p)$  with  $k \leq p/2$ . Now we show that for this function the desired bias lower bound is of order  $k^{-2\beta} (\log k)^2$  as follows,

$$\begin{aligned} \left\| \mathbb{E} \hat{\Sigma}_k^B - \Sigma \right\| &\geq \sup_{x \in [-\pi, \pi]} \left| 2 \sum_{m=k+1}^p \left(1 - \frac{m}{p}\right) \sigma_m \cos mx \right| \\ &\geq 2C \cdot 4^{-t\beta} \cdot \sum_{m=k+1}^{k+4^t} \left(1 - \frac{m}{p}\right) \frac{1}{m-k} \\ &\geq \frac{C}{4} 4^{-t\beta} \log k \geq \frac{C}{4} k^{-\beta} \log k. \quad \blacksquare \end{aligned}$$

For  $\beta \geq 1$  the desired special spectral density exists similarly. We omit the proof for the limit of space.

## 8.2 Proof of Lemma 4

We will modify the Woodroffe and Van Ness's proof (1967) a little to a stronger statement of which our first claim here  $\mathbb{E} \left\| \hat{\Sigma}_k^B - \Sigma \right\|^2 \geq c \frac{k \log k}{np}$  is just a simple consequence. Only a brief proof is given here. For more details, refer to Woodroffe and Van Ness (1967). According to (49), we have that  $\left\| \hat{\Sigma}_k^B - \Sigma \right\| \geq \sup_{\lambda \in [-\pi, \pi]} \left| \tilde{\sigma}_0 - 1 + 2 \sum_{m=1}^{k-1} \left(1 - \frac{m}{p}\right) \tilde{\sigma}_m \cos m\lambda \right|$ . Here we will prove a stronger result, as  $p \rightarrow \infty$

$$\left( \frac{np}{4k \log k} \right)^{1/2} \sup_{\lambda \in [-\pi, \pi]} \left| \tilde{\sigma}_0 - 1 + 2 \sum_{m=1}^{k-1} \left(1 - \frac{m}{p}\right) \tilde{\sigma}_m \cos m\lambda \right| \rightarrow 1 \text{ in probability.}$$

Recall that  $\tilde{\sigma}_m = \frac{1}{p-m} \frac{1}{n} \sum_{l=1}^n \sum_{s-t=m} X_s^{(l)} X_t^{(l)}$  where  $X_s^{(l)}$  are i.i.d. standard normal for all  $l = 1, \dots, n$  and  $s = 1, \dots, p$ . We could write

$$\left(\frac{np}{k}\right)^{1/2} \left( \tilde{\sigma}_0 - 1 + 2 \sum_{m=1}^{k-1} \left(1 - \frac{m}{p}\right) \tilde{\sigma}_m \cos m\lambda \right) = Z_p(\lambda) - r_p(\lambda) + t_p$$

where  $0 \leq \lambda \leq \pi$ , and

$$Z_p(\lambda) = p^{-1/2} \sum_{t=1}^p Z_{p,t}(\lambda) \quad (58)$$

$$Z_{p,t}(\lambda) = 2(kn)^{-1/2} \sum_{l=1}^n \sum_{v=1}^{k-1} X_t^{(l)} X_{t+v}^{(l)} \cos(v\lambda) \quad (59)$$

$$= 2(kn)^{-1/2} \sum_{l=1}^n \sum_{v=1}^{k-1} \omega_{unif}(v/k) X_t^{(l)} X_{t+v}^{(l)} \cos(v\lambda) \quad (60)$$

with  $\omega_{unif}(x) = 1\{|x| < 1\}$ , and

$$\begin{aligned} r_p(\lambda) &= 2(npk)^{-1/2} \sum_{l=1}^n \sum_{t=p-k+2}^p \sum_{v=p-t+1}^{k-1} X_{t+v}^{(l)} X_t^{(l)} \cos(v\lambda) \\ t_p &= (npk)^{-1/2} \sum_{l=1}^n \sum_{t=1}^p \left[ (X_t^{(l)})^2 - 1 \right]. \end{aligned}$$

In the proof of our second claim, we need to replace uniform kernel by another kernel but the remaining part of the proof is similar. Since for  $p = 1, 2, \dots$

$$\mathbb{E} \sup_{\lambda} |r_p(\lambda)| \leq 2(npk)^{-1/2} \sum_{v=1}^{k-1} \mathbb{E} \left| \sum_{l=1}^n \sum_{t=p-v+1}^p X_{t+v}^{(l)} X_t^{(l)} \right|$$

$$\mathbb{E} \left| \sum_{l=1}^n \sum_{t=p-v+1}^p X_{t+v}^{(l)} X_t^{(l)} \right|^2 \leq vn$$

and  $k = O(p^\kappa)$  for some  $\kappa < \frac{2}{5}$ , we have  $\sup_{\lambda} |r_p(\lambda) + t_p| = o_p((\log k)^{-1})$  as  $p \rightarrow \infty$ . Hence it suffices to consider the processes  $Z_p(\lambda)$ ,  $0 \leq \lambda \leq \pi$  defined by (58) and (59). The random variables  $Z_{p,1}(\lambda), Z_{p,2}(\lambda), \dots, Z_{p,p}(\lambda)$ ,  $0 \leq \lambda \leq \pi$ , have the desirable property of  $k$ -dependence, which we will now exploit. Define  $q = q_p = \lfloor k(\log k)^4 \rfloor$  where  $\lfloor \cdot \rfloor$  denotes the greatest integer function. We may write  $p = qd + r$  where  $0 \leq r < q$ . Let us define for  $i = 1, \dots, d$ .

$$\begin{aligned} U_{p,i}(\lambda) &= q^{-1/2} (Z_{p,(i-1)q+1}(\lambda) + \dots + Z_{p,iq-k}(\lambda)) \\ V_{p,i}(\lambda) &= k^{-1/2} (Z_{p,iq-k+1}(\lambda) + \dots + Z_{p,iq}(\lambda)) \\ V_{p,0}(\lambda) &= Z_{p,dq+1}(\lambda) + \dots + Z_{p,p}(\lambda) \\ U_p(\lambda) &= d^{-1/2} \sum_{i=1}^d U_{p,i}(\lambda) \quad \text{and} \quad V_p(\lambda) = d^{-1/2} \sum_{i=1}^d V_{p,i}(\lambda) \end{aligned}$$

Then clearly

$$Z_p(\lambda) = (qd/p)^{1/2} \left( U_p(\lambda) + (k/q)^{1/2} V_p(\lambda) \right) + p^{-1/2} V_{p,0}(\lambda)$$

The proof of  $\max_{0 \leq \lambda \leq \pi} |V_{p,0}(\lambda)| = o(p^{1/2}(\log k)^{-1})$  is similar to the proof for  $\sup_{\lambda} |r_p(\lambda)| = o_p((\log k)^{-1})$ . Next we will truncate  $U_{p,i}(\lambda)$  and  $V_{p,i}(\lambda)$  as follow

$$\begin{aligned} U_{p,i}(\lambda)' &= U_{p,i}(\lambda) \{|U_{p,i}(\lambda)| \leq p^{0.3}\} \\ V_{p,i}(\lambda)' &= V_{p,i}(\lambda) \{|V_{p,i}(\lambda)| \leq p^{0.3}\} \\ U_{p,i}(\lambda)'' &= (U_{p,i}(\lambda)' - EU_{p,i}(\lambda)')/Var(U_{p,i}(\lambda)')^{1/2} \\ V_{p,i}(\lambda)'' &= (V_{p,i}(\lambda)' - EV_{p,i}(\lambda)')/Var(V_{p,i}(\lambda)')^{1/2} \end{aligned}$$

and let  $U_p(\lambda)', U_p(\lambda)'', V_p(\lambda)', V_p(\lambda)''$  be  $d^{-1/2}$  times their respective sums. Note all of them are sums of independent identically distributed random variables.

Before showing  $V_p(\lambda)$  is negligible, we need some lemmas. For the proof of these lemmas, please refer to Woodroffe and Van Ness (1967). The first lemma is a standard result of trigonometric polynomial and the last two are based on Lemma (10) and Lemma (11), which are not hard to prove.

**Lemma 9** *Let  $p(\lambda) = \sum_{v=-k}^k \alpha_v \exp(iv\lambda)$  be a trigonometric polynomial. Define  $\lambda_j = \pi(j/rk)$ ,  $|j| \leq rk$ . Then*

$$\max_{|\lambda| \leq \pi} p(\lambda) \leq \max_{|j| \leq rk} |p(\lambda_j)/(1 - 3\pi r^{-1})|$$

**Lemma 10** *The random variables  $Z_{p,1}(\lambda), \dots, Z_{p,p}(\lambda)$  have zero means and covariance*

$$Cov(Z_{p,1}(\lambda_1), Z_{p,1}(\lambda_2)) = (4/k) \sum_{v=1}^{k-1} (\cos v\lambda_2)(\cos v\lambda_1)$$

*If  $t_1 < t_2 < t_3 < t_4$  and  $0 \leq \lambda_i \leq \pi$ , then*

$$\mathbb{E}(Z_{p,t_1}(\lambda_1)Z_{p,t_2}(\lambda_2)) = 0 = \mathbb{E}(\prod_{i=1}^4 Z_{p,t_i}(\lambda_i))$$

*Moreover, there exists a constant  $C$  for which*

$$\begin{aligned} \left| \mathbb{E}(\prod_{i=1}^4 Z_{p,t_i}(\lambda_1)) \right| &\leq C \quad \text{if } t_1 = t_2 \text{ and } t_3 = t_4 \\ &\leq C(nk)^{-1} \quad \text{if } t_1 = t_2 \neq t_3 \neq t_4 \end{aligned}$$

**Lemma 11** *Let  $h(p) = k\lambda_p$  and  $0 \leq \lambda_p < \pi$ . If  $h(p) \rightarrow \infty$  as  $p \rightarrow \infty$ , then  $(2/k) \sum_{v=0}^{k-1} \cos v\lambda_p = O(h(p)^{-1})$  as  $p \rightarrow \infty$ ; if  $\liminf h(p) \geq 1$ , then  $\overline{\lim}(2/k) \left| \sum_{v=0}^{k-1} \cos v\lambda_p \right| < \|\omega_{unif}\|_2^2 = 2$ .*



**Lemma 12** Let  $\sigma_p^2(\lambda) = \text{Var}(Z_{p,1}(\lambda))$ , then  $\sigma_p^2(\lambda)$  is uniformly bounded and

$$\sigma_p^2(\lambda) \rightarrow 2 \quad \text{as } p \rightarrow \infty$$

uniformly on  $[k^{-1} \log k, \pi]$

**Lemma 13** Let  $\gamma_p(\lambda_1, \lambda_2)$  be the correlation of  $Z_{p,1}(\lambda_1)$  and  $Z_{p,1}(\lambda_2)$ ,  $0 \leq \lambda_i \leq \pi$ , then

$$\begin{aligned} \sup_{|\lambda_1 - \lambda_2| \geq (\log k)^2 k^{-1}} |\gamma_p(\lambda_1, \lambda_2)| &= O((\log k)^{-2}) \\ \overline{\lim}_{p \rightarrow \infty} \sup_{|\lambda_1 - \lambda_2| \geq k^{-1}} |\gamma_p(\lambda_1, \lambda_2)| &< 1 \end{aligned}$$

Based on Lemma (10), it's not hard to see  $\mathbb{E}|V_{p,i}(\lambda)|^4 \leq C$  and  $E|U_{p,i}(\lambda)|^4 \leq Cq(nk)^{-1}$ . According to Lemma (9), the fact that  $V_p(\lambda)$  is negligible follows from

$$P(V_p(\lambda_{p,j})' \neq V_p(\lambda_{p,j}), \text{ for some } j) \rightarrow 0 \quad (61)$$

$$\max_j |V_p(\lambda_{p,j})' - V_p(\lambda_{p,j})''| \leq O(1) \max_j |V_p(\lambda_{p,j})'| + o(1) \quad (62)$$

$$\max_j |V_p(\lambda_{p,j})''| = o_p(\log k) \quad (63)$$

as  $p \rightarrow \infty$  where  $\lambda_{p,j} = \pi j / \lfloor k \log k \rfloor$ ,  $j = 0, \dots, \lfloor k \log k \rfloor$ . Equation (61) follows from  $\mathbb{E}|V_{p,i}(\lambda)|^4 \leq C$  since

$$P(V_p(\lambda_{p,j})' \neq V_p(\lambda_{p,j}), \text{ for some } j) \leq \sum_i \sum_j p^{-6/5} \mathbb{E}|V_{p,i}(\lambda_{p,j})|^4 \leq Cp^{-1/5}.$$

Equation (62) can be shown to follow similarly. Since for  $\varepsilon > 0$ ,

$$P(\max_j |V_p(\lambda_{p,j})''| \geq \varepsilon \log k) \leq \sum_j P(|V_p(\lambda_{p,j})''| \geq 2(2 \log k)^{1/2})$$

when  $p$  is sufficiently large, Equation (63) is then an easy consequence of the first part of Lemma (14) below.

**Lemma 14** Let  $\Phi(\cdot)$  denote the standard univariate normal distribution function and  $\varphi_r(\cdot, \cdot)$  to denote the standard bivariate normal density with correlation  $r$ . If  $0 < z_p \rightarrow \infty$  and  $z_p = o(\log k)$  as  $p \rightarrow \infty$ , then as  $p \rightarrow \infty$

$$\begin{aligned} P(|V_p(\lambda)|'') &\geq z_p \sim 2(1 - \Phi(z_p)) \quad \text{as } p \rightarrow \infty \text{ uniformly on } [0, \pi] \\ P(\pm U_p(\lambda_1)'' &\geq z_p, \pm U_p(\lambda_2)'' \geq z_p) \\ &\sim \int_{z_p}^{\infty} \int_{z_p}^{\infty} \varphi_{r_p(\lambda_1, \lambda_2)}(\pm y_1, \pm y_2) dy_1 dy_2 \quad \text{uniformly on } S_p \\ P(\pm U_p(\lambda_i)'' &\geq z_p, i = 1, \dots, v) \sim (1 - \Phi(z_p))^v \quad \text{uniformly on } S_{p,v} \\ P(\pm U_p(\lambda_i)'' &\geq z_p, i = 1, \dots, v) \\ &\sim (1 - \Phi(z_p))^{v-2} \int_{z_p}^{\infty} \int_{z_p}^{\infty} \varphi_{r_p(\lambda_1, \lambda_2)}(\pm y_1, \pm y_2) dy_1 dy_2 \quad \text{uniformly on } S'_{p,v} \end{aligned}$$

where  $S_p = \{(\lambda_1, \lambda_2), 0 \leq \lambda_i \leq \pi \text{ and } |\lambda_1 - \lambda_2| \geq k^{-1}\}$ ,  $S_{p,v} = \{(\lambda_1, \dots, \lambda_v), 0 \leq \lambda_i \leq \pi \text{ and } \min_{i \neq j} |\lambda_i - \lambda_j| \geq k^{-1}(\log k)^2\}$  and  $S'_{p,v} = \{(\lambda_1, \dots, \lambda_v), \lambda_2 - \lambda_1 \geq k^{-1} \text{ and } \lambda_i - \lambda_{i-1} \geq k^{-1}(\log k)^2 \text{ } i = 3, \dots, v\}$ .

As a consequence, we have  $P(|U_p(\lambda)''| \geq z_p) \sim 2(1 - \Phi(z_p))$  as  $p \rightarrow \infty$  uniformly on  $[0, \pi]$ . The result of Lemma (12) is also true for  $\sigma_p^2(\lambda)' = \text{Var}(U_{p,i}(\lambda)')$ . i.e. uniformly bounded and

$$\sigma_p^2(\lambda)' \rightarrow 2 \text{ as } p \rightarrow \infty$$

uniformly on  $[k^{-1} \log k, \pi]$ .

Please refer to Woodroffe and Van Ness (1967) for full details. Basically lemma (13) and the truncation are used to prove this lemma.

Now it's enough to show

$$\left(\frac{1}{4 \log k}\right)^{1/2} \sup_{\lambda \in [-\pi, \pi]} U_p(\lambda) \rightarrow 1 \text{ in probability}$$

To further simplify it, we note that  $P(U_p(\lambda_{p,j})' \neq U_p(\lambda_{p,j}), \text{ for some } j) \rightarrow 0$  by the same argument (61) above for  $V_p(\lambda_{p,j})$ . By the fact  $\mathbb{E} |U_{p,i}(\lambda)|^4 \leq Cq(nk)^{-1}$  and Lemma (10) it's easy to see that

$$(U_p(\lambda)'' \sigma_p(\lambda)' - U_p(\lambda)') = d^{1/2} (EU_{p,i}(\lambda)') = o(1)$$

According to Lemma (9), it's enough to show

$$\lim P(\max_j |U_p(\lambda_{p,j})''| \sigma_p(\lambda_{p,j})') \geq (1 + \varepsilon) \sqrt{2} (2 \log k)^{1/2}) = 0, \quad (64)$$

$$\lim P(\max_j |U_p(\lambda_{p,j})''| \sigma_p(\lambda_{p,j})') \leq (1 - \varepsilon) \sqrt{2} (2 \log k)^{1/2}) = 0 \quad (65)$$

To establish (64), let  $S$  be the set of integers  $j$  for which  $1 \leq j \leq \lfloor k \log k \rfloor$  and  $\lambda_{p,j} \geq k^{-1} \log k$ . Then if  $\varepsilon = 2\varepsilon'$  is given, we find from last two parts of Lemma (14) that for  $p$  sufficiently large

$$\begin{aligned} P(\max_{j \in S} |U_p(\lambda_{p,j})''| \sigma_p(\lambda_{p,j})' &\geq (1 + \varepsilon) \sqrt{2} (2 \log k)^{1/2}) \\ &\leq \sum_{j \in S} P(|U_p(\lambda_{p,j})''| \geq (1 + \varepsilon') (2 \log k)^{1/2}) \\ &\leq 4k \log k (1 - \Phi((1 + \varepsilon') (2 \log k)^{1/2})) = o(1) \end{aligned}$$

and

$$\begin{aligned}
P(\max_{j \notin S} |U_p(\lambda_{p,j})''| \sigma_p(\lambda_{p,j})' &\geq (1 + \varepsilon) \sqrt{2} (2 \log k)^{1/2}) \\
&\leq \sum_{j \notin S} P(|U_p(\lambda_{p,j})''| \geq c(2 \log k)^{1/2}) \\
&\leq 2(\log k)^2 (1 - \Phi(c(2 \log k)^{1/2})) = o(1)
\end{aligned}$$

as  $p \rightarrow \infty$ , where  $c^2 > 0$  is a lower bound for  $2/\sigma_p^2(\lambda)'$ . This establishes (64). (65) could be established using Lemma (14). Full details are given in the technical report of Woodroffe and Van Ness (1967). Therefore we finish the first part of Lemma 4.

To show the second claim,  $\mathbb{E} \left\| \hat{\Sigma}_k^B - \Sigma \right\|^2 \geq c \frac{p^\kappa \log p}{np}$  for  $k \geq p^\kappa$ , first we set  $k_0 = p^\kappa$  and note that

$$\left\| \hat{\Sigma}_k^B - \Sigma \right\| \geq \left\| \left( \hat{\Sigma}_k^B - \Sigma \right)_{k_0 \times k_0} \right\| \geq \sup_{\lambda \in [-\pi, \pi]} \left| \tilde{\sigma}_0 - 1 + 2 \sum_{m=1}^{k_0-1} \left( 1 - \frac{m}{k_0} \right) \tilde{\sigma}_m \cos m\lambda \right|,$$

where  $\left( \hat{\Sigma}_k^B - \Sigma \right)_{k_0 \times k_0}$  is the upper  $k_0 \times k_0$  submatrix of  $\hat{\Sigma}_k^B - \Sigma$ . Note the subtle difference between the right hand side of the inequality above and that in the (60) at the beginning of this proof. This time we need to replace the uniform kernel in the expression of  $Z_{p,t}(\lambda)$  by another kernel.  $Z_{p,t}(\lambda)$  can be written as

$$Z_{p,t}(\lambda) = 2(k_0 n)^{-1/2} \sum_{l=1}^n \sum_{v=1}^{k_0-1} \omega_p(v/k_0) X_t^{(l)} X_{t+v}^{(l)} \cos(v\lambda)$$

where kernel  $\omega_p(x) = 1\{|x| < 1\} \frac{1-|x|}{1-|x|p^\kappa-1}$ . The proof is pretty similar to that of the first claim if we could prove the following fact corresponding to Lemma (11).

**Lemma 15** *Let  $h(p) = k_0 \lambda_p$  and  $0 \leq \lambda_p < \pi$ . If  $h(p) \rightarrow \infty$  as  $p \rightarrow \infty$ , then*

$$(2/k_0) \sum_{v=0}^{k_0-1} \omega_p(v/k_0)^2 \cos v\lambda_p = O(h(p)^{-1}) \quad \text{as } p \rightarrow \infty$$

If  $\underline{\lim} h(p) \geq 1$ , then

$$\overline{\lim}(2/k_0) \left| \sum_{v=0}^{k-1} \omega_p(v/k_0)^2 \cos v\lambda_p \right| < \|\omega_{triangle}\|_2^2 = 2/3$$

where  $\omega_{triangle}(x) = 1\{|x| < 1\}(1 - |x|)$ .

Once this is proved, there is no difficulty in showing the desired result, following the steps in the above proof (refer to Woodroffe and Van Ness (1967) for details)

$$\left( \frac{np}{2 \|\omega_{triangle}\|_2^2 k_0 \log k_0} \right)^{1/2} \sup_{\lambda \in [-\pi, \pi]} \left| \tilde{\sigma}_0 - 1 + 2 \sum_{m=1}^{k_0-1} \left( 1 - \frac{m}{k_0} \right) \tilde{\sigma}_m \cos m\lambda \right| \rightarrow 1$$

in probability. Finally note that  $E \left\| \hat{\Sigma}_k^B - \Sigma \right\|^2 \geq c \frac{k_0 \log k_0}{np}$  is just a consequence of it. The Lemma (15) cannot be shown directly using the method in Woodroffe and Van Ness (1967) since the requirement of the kernel in that paper is not satisfied by our kernel, namely that its second derivative at 0 doesn't exist. Hence we prove it in details. First of all notice that to prove Lemma (15) we only need to show

$$(2/k_0) \sum_{v=0}^{k_0-1} \omega_{triangle}(v/k_0)^2 \cos v\lambda_p = O(h(p)^{-1}) \quad \text{as } p \rightarrow \infty \quad (66)$$

and  $\overline{\lim}(2/k_0) \left| \sum_{v=0}^{k-1} \omega_{triangle}(v/k_0)^2 \cos v\lambda_p \right| < \|\omega_{triangle}\|_2^2 = 2/3$  because it is easy to see that

$$(2/k_0) \sum_{v=0}^{k_0-1} |\omega_{triangle}(v/k_0)^2 - \omega_p(v/k_0)^2| = O(p^{\kappa-1}) = o(h(p)^{-1})$$

noting the facts  $h(p) = k_0 \lambda_p = O(p^\kappa)$  and  $\kappa < \frac{2}{5}$ . The proof of

$$\overline{\lim}(2/k_0) \left| \sum_{v=0}^{k-1} \omega_{triangle}(v/k_0)^2 \cos v\lambda_p \right| < \|\omega_{triangle}\|_2^2 = 2/3$$

is straightforward (cf. Woodroffe and Van Ness (1967) page 1562) and here we only focus on the proof of the fact (66). Let  $W(y)$  be the Fourier transform of the kernel  $\omega_{triangle}(x)$ . i.e.

$$W(y) = (2\pi)^{-1} \int e^{-ixy} \omega_{triangle}(x) dx = \pi^{-1} \frac{1 - \cos x}{x^2} \geq 0$$

Routine Fourier analysis yields

$$k_0^{-1} \sum_{v=0}^{k_0-1} \omega_{triangle}(v/k_0)^2 \cos v\lambda_p = \int_{-\pi}^{\pi} \frac{\sin[(k_0-1)(y+\lambda_p)]}{\sin[2^{-1}(y+\lambda_p)]} W_p(y) dy$$

where  $W_p(y) = \sum_{m=-\infty}^{\infty} W * W(k_0 y + 2mk_0 y)$ ,  $*$  denotes convolution in  $L_1(R)$  space. Notice  $W_p(y)$  is non-negative and  $\lambda_p \leq \pi$ , we have

$$\begin{aligned} & \int_{2|y| \leq \lambda_p} \left| \frac{\sin[(k_0-1)(y+\lambda_p)]}{\sin[2^{-1}(y+\lambda_p)]} \right| W_p(y) dy \\ & \leq |\sin(\lambda_p/4)|^{-1} \int_{-\pi}^{\pi} W_p(y) dy = \frac{1}{|\sin(\lambda_p/4)| k_0} \int_{-\infty}^{\infty} W * W(y) dy \\ & = \frac{2\pi}{|\sin(\lambda_p/4)| k_0} \omega_{triangle}^2(0) = O(h(p)^{-1}) \end{aligned}$$

Notice that  $W * W(y) = \frac{2(\lambda - \sin \lambda)}{\pi \lambda^3}$ , we have

$$\begin{aligned} \int_{2|y| > \lambda_p} \left| \frac{\sin[(k_0-1)(y+\lambda_p)]}{\sin[2^{-1}(y+\lambda_p)]} \right| W_p(y) dy & \leq k_0 \int_{2|y| > h(p)/k_0} W_p(y) dy \leq \int_{2|y| > h(p)} W * W(y) dy \\ & \leq C \int_{h(p)/2}^{\infty} y^{-2} dy = O(h(p)^{-1}) \end{aligned}$$

The two inequalities above show the desired result and hence Lemma (15) is proved.  $\blacksquare$